MBMG
Metabarcoding & Metagenomics

Metabarcoding and Metagenomics 10: 69–104 (2026)
DOI: 10.3897/mbmg.10.173634

Research Article

Modular universal *rbc*L primers for (meta)barcoding: a systematic case study clarifying the identity of the marine macroalga *Corallina armata* (Rhodophyta, Corallinales)

Aimee L. van der Reis[1], Mary A. Sewell[1], Rachael M. Wade[2], Wendy A. Nelson[1,3]

1 *School of Biological Sciences, University of Auckland, Auckland, New Zealand*
2 *Department of Biological Sciences, California State Polytechnic University, Arcata, California, USA*
3 *Auckland War Memorial Museum Tāmaki Paenga Hira, Auckland, New Zealand*

Corresponding author: Aimee L. van der Reis (aimee.vdreis@auckland.ac.nz)

OPEN ACCESS

Academic editor: Hugo de Boer
Received: 2 October 2025
Accepted: 3 January 2026
Published: 6 February 2026

Citation: van der Reis AL, Sewell MA, Wade RM, Nelson WA (2026) Modular universal *rbc*L primers for (meta)barcoding: a systematic case study clarifying the identity of the marine macroalga *Corallina armata* (Rhodophyta, Corallinales). Metabarcoding and Metagenomics 10: e173634. https://doi.org/10.3897/mbmg.10.173634

Copyright: © Aimee L van der Reis et al. This is an open access article distributed under terms of the Creative Commons Attribution License (Attribution 4.0 International – CC BY 4.0).

Abstract

The geniculate coralline algal genus *Corallina* L. (Rhodophyta, Corallinales) presents longstanding challenges in taxonomy and nomenclature. Recent studies have uncovered previously unrecognised diversity within the genus, highlighting the importance of revisiting historical names and type specimens. In this study, we resolve the taxonomic identity of *Corallina armata* through DNA sequencing of the *rbc*L and COI gene regions and designate an epitype to stabilise its nomenclature. Phylogenetic analyses consistently grouped *Corallina* sequences into species-specific clades across both markers. Using population genetic approaches, single nucleotide polymorphism data revealed clear genetic separation of *C. armata* from related species, supported by hierarchical clustering, based on genetic distance and principal component analysis of allele frequencies. A pairwise permutational analysis of variance confirmed significant genetic differentiation amongst species ($p < 0.05$). To support broader taxonomic efforts and utilised in this study, we developed modular universal *rbc*L primer suites for macroalgae, which performed well *in silico* across the GenBank database, indicating strong potential for international application (inclusive of some microalgae). Taxonomic identification results from the short-read sequences for *rbc*L (Ochrophyta and Rhodophyta) and COI showed both primer sets enabled species-level identification, although the *rbc*L primer sets captured less genetic variation than COI. Nonetheless, each gene revealed distinct interspecific relationships, underscoring the value of multi-locus approaches for robust species delimitation in future barcoding and metabarcoding studies.

Key words: Algae, COI, DNA metabarcoding, epitypification, modular universal primer suites, primer development, *rbc*L

Introduction

The geniculate coralline algal genus *Corallina* L. (Rhodophyta, Corallinales), is challenging in terms of its taxonomy and nomenclatural history. In particular, the type species, *C. officinalis* L., once regarded as widely distributed globally, is now considered to be restricted to the North Atlantic and north-eastern Pacific, based on DNA sequence data (Hind et al. 2014; Calderon et al. 2021).

Recent investigations have discovered previously unrecognised diversity in the genus and emphasised the need to examine early names along with type specimens, including some that had been reduced to synonymy with *C. officinalis* (Calderon et al. 2021; Schipper et al. 2023; Wade et al. 2023; Nelson et al. 2025). Morphoanatomy has been shown to be unreliable for distinguishing species of *Corallina*: Calderon et al. (2021) noted that there are no diagnostic morpho-anatomical characters for species in this genus. Type material of several *Corallina* species have been successfully sequenced, which is an essential step to enable clarification of species concepts and the correct application of names (e.g. Hind and Saunders (2013); Hind et al. (2014); Brodie et al. (2021); Calderon et al. (2021); Schipper et al. (2023); Wade et al. (2023); Nelson et al. (2025); Wade et al. (2025)).

A gene region commonly targeted for plant species identification is ribulose-1,5-bisphosphate carboxylase (*rbc*L) gene and is one of two core barcode gene regions in the chloroplast genome proposed by the Consortium for the Barcode of Life (CBOL) Plant working group (CBOL et al. 2009; Kress and Erickson 2012). The *rbc*L gene region tends to be readily amplified across phylogenetically divergent lineages in comparison to some other gene regions (CBOL et al. 2009) and had already been adopted for use in algal species more than a decade before its proposal by CBOL (e.g. Freshwater et al. (1994); Nozaki et al. (1995); Daugbjerg and Andersen (1997)). The ease of amplification and usefulness of *rbc*L has resulted in *rbc*L-rich databases (e.g. the National Center of Biotechnology Information's GenBank). However, there are two caveats that hinder their usefulness. The first is that universal *rbc*L primers are deficient for (meta)barcoding purposes and thereby restrict research (e.g. algae biodiversity assessments). The second caveat is that, despite best efforts, the databases still pose problems, such as missing critical metadata, sequencing errors and taxonomic mislabelling (see Keck et al. (2023)). Taxonomic mislabelling of submitted sequences is one of the more serious problems, because the end-user is typically unaware of the initial misidentifications. Initial misidentification may occur when a specimen has not been identified or verified by an expert or in cases where the systematics of the target organisms is poorly understood. Some genera/species are particularly difficult or, in fact unable, to be identified using solely morphological features, even by experts (e.g. Calderon et al. (2021); Wade et al. (2025)). Thus, sequences from single species may be present in DNA databases with a diversity of assigned names, highlighting the importance of sequencing type materials and incorporating them as references in subsequent analyses.

In New Zealand, the name *C. officinalis* was used almost exclusively for species of *Corallina* for more than a century, with only occasional reference to two other species based on New Zealand type material that were described in the 1840s (*C. armata* Hook f. et Harv., *C. hombronii* (Mont.) Mont. ex Kützing). Sequence data obtained from type material of *C. hombronii* has enabled the correct application of this name, resulting in the synonymising of *C. chamberlainiæ* J.Brodie & R.Mrowicki (Nelson et al. 2025). The identity of *C. armata* (Harvey 1849) has remained unclear since it was first described; the name is not in common use and the need to clarify its status has been evident for some time (Adams 1994; Woelkerling and Nelson 2004). Colenso sent *Corallina* specimens from two collections to W.J. Hooker in London: Colenso 224 ("on tidal

rocks near Mataikona") in July 1844 and Colenso 654 ("Takamaitua bet'n Cape Turnagain & Castlepoint just above low water mk., near 'Black Head") in July 1846. Both Colenso collection numbers are cited in the description of *C. armata* – "New Zealand, Rev. W. Colenso, No. 224 and 654. (v. s. in Herb. T.C.D. comm. cl. Hooker)". There are specimens with these collection numbers housed in the Herbaria of Trinity College, Dublin and the Natural History Museum, London (TCD and BM, respectively, Thiers (2025)) and they constitute isosyntypes. The Colenso specimen 654 (TCD0011847) was annotated as 'Type' by H.W. Johansen (dated 22 September 1967), but this designation has not been validly published according to the Code (Turland et al. 2025). The morphological attributes of the *C. armata* isosyntype material and the accompanying illustrations provided with the type descriptions are not sufficient to enable unequivocal identification of *C. armata*. Attempts were made to sequence the material of *C. armata* held in TCD and in BM. Both Sanger sequencing and short-read high throughput sequencing were attempted, but neither method returned any usable organellar sequences, suggesting that the DNA has become too degraded.

Here, we have two aims: first is to resolve the identity of *C. armata*, with the ultimate goal of designating an epitype to stabilise the species nomenclature and facilitate future *Corallina* phylogenetic research. Multiple markers are often capitalised to generate robust results (e.g. Evans et al. (2007); Freshwater et al. (2010); Kezlya et al. (2023); Borer et al. (2025); San et al. (2025)). Thus, we employed both *rbc*L and COI as historically targeted loci to explore phylogenetic species limits for the *C. armata* clade. We explored the use of data analysis methods typically used in population genetic studies and how they might facilitate taxonomic studies in the genus *Corallina* (and other systematic studies in general). The second aim, instrumental to achieving the first, was to design and develop modular universal *rbc*L primer suites for Chlorophyta, Ochrophyta and Rhodophyta – with some of the designed Ochrophyta and Rhodophyta primer sets being used in this study. While the initial aim of primer development was to assist with the DNA amplification of New Zealand macroalgal species to facilitate *rbc*L reference sequence database entries for future (meta)barcoding studies, their utility was further tested *in silico* against the current extensive entries for Chlorophyta, Ochrophyta and Rhodophyta in GenBank to determine their universality on a global scale. We provide evidence that the *rbc*L primers developed will assist the phycological community with difficult taxonomic questions and ecologists through application of these primers in DNA metabarcoding studies – helping to bridge the gap between barcoding and metabarcoding research.

Materials and methods

The method section is presented in two parts. The first section focuses on *Corallina armata* and resolving its taxonomic identity, the main aim of the study. The second section provides the necessary details of the method for primer design and development. The data generation and analyses largely utilised Bourne Again Shell (Bash) as a command interpreter and programming language and, for statistical computing, the R v.4.1.1 (RStudio v.2024.09.0, Build 375; R Core Team (2024)) programming language was used with heavy reliance on ggplot2 v.3.5.2 (Wickham 2016) for graphics. To note, Heterokontophyta is not yet recognised in GenBank and thus taxa that belong to the phylum Heterokontophyta

are labelled under Ochrophyta (Guiry et al. 2023; Guiry 2024). To this end, Ochrophyta has been used consistently throughout the manuscript due to the reliance on GenBank and to aid reproducibility.

Corallina armata case study

Specimen collection

The collections targeted the location descriptions of Colenso, the morphological attributes of the *C. armata* isosyntype material and the accompanying illustrations provided with the type descriptions. Fresh specimens of material considered to conform to *C. armata* (n = 4) and *C. berteroi* (n = 1) Montagne ex Kützing were collected in 2024 from Mataikona tidal pools, Wairarapa, southern North Island (New Zealand) (Fig. 1; Suppl. material 1: table S1). While field

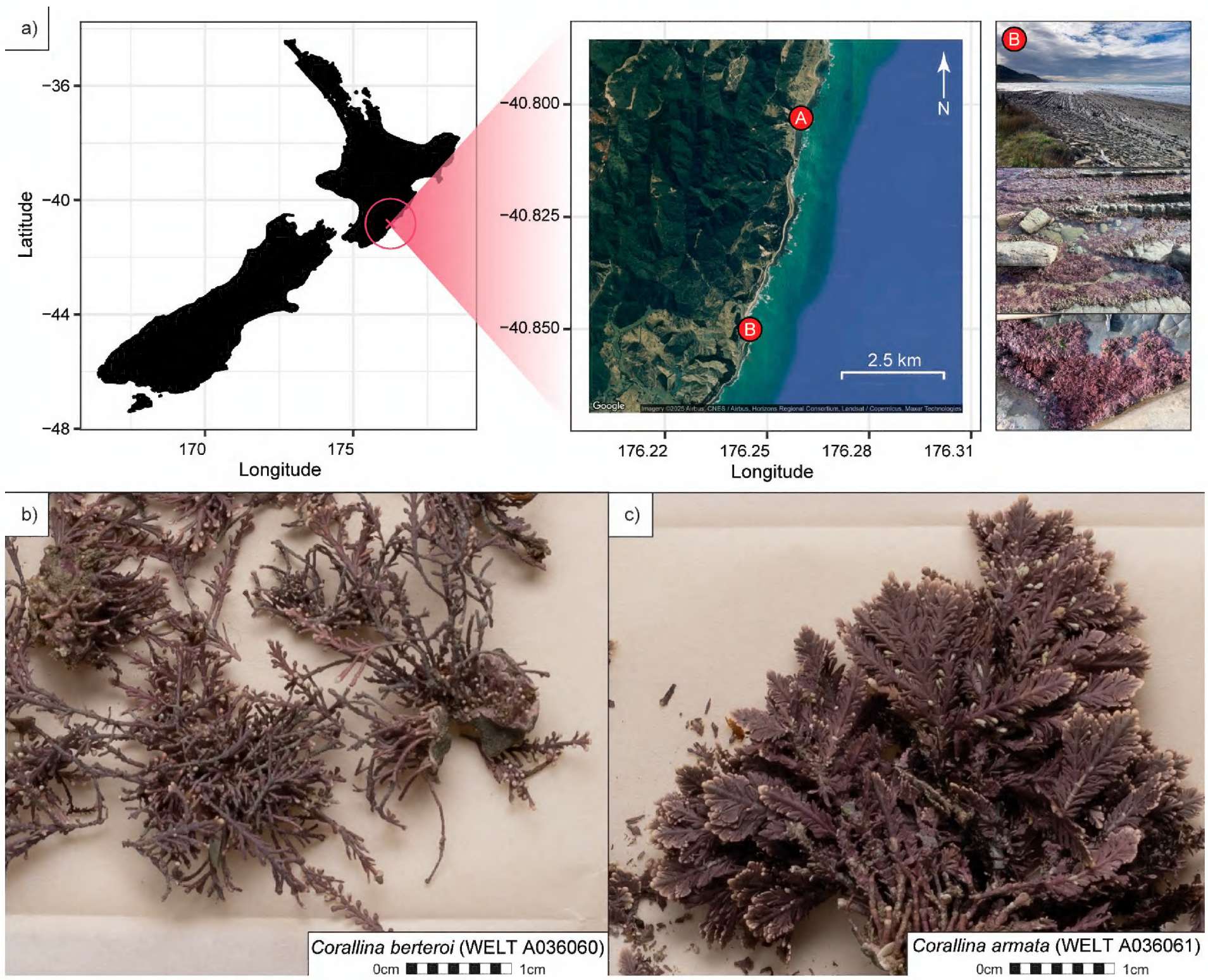

Figure 1. *Corallina* collection locations and photographs of exemplar material. **a)** The left map shows the approximate location of collection at Mataikona, Wairarapa on the east coast of North Island (New Zealand). The right map shows the satellite image and collection location detailed (A – WELT A036058-59; B – WELT A036060-62; Suppl. material 1: table S1). The adjacent images are of collection location B and *Corallina armata*. The maps are produced using ggmap (v.4.0.1; Kahle and Wickham (2013)) and ggplot2 (v.3.5.2; Wickham (2016)); **b)** *Corallina berteroi* (Cber1); **c)** *Corallina armata* (Carm3) collected from site B were used in this study (images CC-BY 4.0, Te Papa).

identification of *Corallina* species in New Zealand is considered unreliable, the specimens were placed into these two species based on external morphology, as *C. armata* (Carm1–4) and *C. berteroi* (Cber1) (later confirmed by sequencing data). The samples were air dried, with subsamples preserved in silica and the specimens deposited in the Herbarium of the Museum of New Zealand Te Papa Tongarewa (WELT, Thiers (2025)).

DNA extraction, amplification and sequencing

DNA was extracted from the *Corallina* specimens using the Nucleospin Tissue DNA extraction kit (Macherey-Nagel), following the manufacturer's instructions. Specimens were chopped into small pieces using sterile scalpel blades prior to overnight lysis. The quality and quantity of the DNA was checked on a 0.8% agarose gel and visualised using Gel Red (Biotium), in a Gel Doc XR+ (Bio-Rad).

For short-read metabarcoding, the DNA was amplified for the selected *rbc*L primer pairs for both Ochrophyta and Rhodophyta (given the possibility of *Corallina* amplification from either; see section 'Primer design, pairing and testing') and COI (Geller et al. (2013); Leray et al. (2013); specifically included for comparison of 'mini barcode' species-level identification as similar in length to *rbc*L short-read primer pairs) (Table 1). The polymerase chain reactions (PCRs)

Table 1. Primer pairs, gene targeted, amplicon size and PCR protocols followed. Two gene regions (COI and *rbc*L) were targeted to amplify *Corallina armata* and *C. berteroi* specimens. Next-generation sequencing: For amplification of a 313 base pair (bp) region of COI, the 'Leray' primers were used (Geller et al. 2013; Leray et al. 2013). For *rbc*L, targeting less than 350 bp regions, the modular primer suites (this study – see section 'Primer design, pairing and testing') were used; the Rhodophyta primers are denoted as 'Rho' and the Ochrophyta as 'Och'. Nextera adapters were added to the 5' end of the primers (Illumina 2013). For Sanger sequencing: Rho1 and Rho8 were used to capitalise on amplifying the longest portion of *rbc*L that would be possible with the selected Rho primers.

Gene	Primer set	Size (bp)	Primer 5'-3'	PCR profile
COI	Leray	313	F: GGWACWGGWTGAACWGTWTAYCCYCC	94 °C 1'
			R: TANACYTCNGGRTGNCCRAARAAYCA	35 × [94 °C 15", 54 °C 90", 72 °C 10"]
				75 °C 5'
*rbc*L 'Rho'	Rho1-3	325	1: GCWAARATGGGHTAYTGGGAT	94 °C 1'
			3: GGHADRCGCATRTCTTCWA	35 × [94 °C 15", 54 °C 30", 72 °C 10"]
	Rho5-7	299	5: CARCCWTTYATGCGYTGGAR	72 °C 5'
			7: DCCWGCCATDCKCATCCAYT	
	Rho7-8	273	7: ARTGGATGMGHATGGCWGGH	
			8: TTWGCWGTWGCWCCWGCTTG	
*rbc*L 'Och'	Och1-3	281	1: TGGAYDGTHGTWTGGA	94 °C 1'
			3: CCTARKAAWGGACGDCMAAA	35 × [94 °C 15", 50 °C 30", 72 °C 10"]
	Och3-6	316	3: TTTKGHCGTCCWTTMYTAGG	72 °C 5'
			6: GCCCADATWGCCATHSTTTG	
	Och6-9	298	6: CAAASDATGGCWATHTGGGC	
			9: ARTTSRTGCATTTGRCCACA	
*rbc*L	Rho1-8	1,158	1: GCWAARATGGGHTAYTGGGAT	94 °C 1'
			8: TTWGCWGTWGCWCCWGCTTG	35 × [94 °C 15", 54 °C 30", 72 °C 1']
				72 °C 5'

were performed using 12.5 µl MyTaq Red Mix (Bioline; Meridian Bioscience), 0.5 µl of each primer (10 µM), 10.5 µl UltraPure DNase/RNase-Free Distilled Water (Invitrogen; Thermo Fisher Scientific) and 1 µl DNA. Negative controls were included in the DNA extractions (extraction blank – no tissue added) and subsequently in the PCRs, to check for possible contamination. Possible contamination was also monitored by including a blank in every PCR run (PCR blank – no template added). A temperature gradient was initially used to optimise the annealing temperatures for the Ochrophyta and Rhodophyta primers.

The short-read PCR products were pooled by gene (primer suites pooled by phylum targeted, i.e. 'Rho' and 'Och'; Table 1), by specimen, before proceeding with DNA clean-up using Agencourt AMPure XP (Beckman Coulter) following the Illumina '16S Metagenomic Sequencing Library Preparation' protocol for PCR clean-up (Illumina 2013). The concentrations of the purified PCR products were determined using a Qubit dsDNA HS Assay Kit (Invitrogen) following the manufacturer's instructions and brought to equal concentration before indexing for multiplexed sequencing. Sequencing was done on an Illumina Miseq platform (2 × 300 pair-end; V3 chemistry; shared run) through Auckland Genomics (University of Auckland, New Zealand).

Sanger sequencing (using Rho1, i.e. forward direction only) was performed for every specimen (Carm1-4, Cber1) using the Rho1-8 primer pair for initial species identifications and as a reference for Ochrophyta and Rhodophyta short-read sequence comparisons. Clean-up of the PCR products was conducted as above. Sequencing reactions to incorporate the fluorescent dyes were performed using the ABI PRISM BigDye Terminator Sequencing Kits (v.3.1; Applied Biosystems; Thermo Fisher Scientific). Agencourt CleanSEQ magnetic beads (Beckman Coulter) were used to remove the unincorporated fluorescent dyes. Sequencing was undertaken using an ABI PRISM 3500XL Genetic Analyzer using POP-7 polymer and a 50 cm capillary array sequencer by Auckland Genomics. Raw sequencing data were processed using the Sequencing Analysis Software (v.7; Thermo Fisher Scientific) and the resulting data generated were then put through Auto PeakTrace 6 RP (Nucleics) software to improve the quality of sequences. Carm2 resulted in a noisy sequence and, therefore, it was re-sequenced (a new PCR product used); 5% dimethyl sulphoxide was included in an attempt to improve sequencing quality (Winship 1989). The sequences were viewed in Geneious (Geneious Prime v.2025.0.2; Build 2024-10-30) and trimmed to remove poor quality bases at both ends.

Sequence processing

Cutadapt (v.4.1; Martin (2011)) was used to trim the forward and reverse primers from the short-read sequences resulting from the multiplexed, high throughput sequencing. Primers were removed if an exact sequence match could be found anchored at the beginning of the sequence and no indels or errors were allowed in the primer sequence. All untrimmed sequences were discarded. Qiime 2 (v.2023.5; Bolyen et al. (2019)) was used to visualise the initial sequence quality (*demux summarize*). DADA2 (within Qiime 2; Callahan et al. (2016)) was used for sequence processing, based on quality scores (sequences truncated to keep a median quality score greater than 35), denoising, merging and chimera formation, ensuring only high quality paired-end

sequences were retained. Minimum sequence lengths (Rho = 265 base pairs [bp], Och = 265 bp, Leray = 270 bp) were set to further ensure no spurious alignments would be retained. The top 10 taxonomic assignments (MegaBLAST v.2.16.0; nucleotide database 2024-10; Morgulis et al. (2008)) for each amplicon sequence variants (ASVs; Callahan et al. (2017)) were inspected using custom code. Briefly, each assignment required sequence coverage of at least 95% and an e-value of 0.001 or less. Thereafter, the percentage identity was used to identify the best possible match; however, if more than one of the hits had the same percentage identity then the taxonomic assignment used the lowest common ancestor (e.g. if different species of the same genus have the same percentage identity, then genus level is assigned). All negatives were filtered out during the quality control process. For conservative community analyses, the relative read abundance (RRA) of each ASVs, per sample, was calculated and retained if there were at least 10 reads and the RRA was > 1%. These filtering parameters were not implemented when extracting *Corallina* sequences for phylogenetic analyses, given the bias of amplifying Ochrophyta taxa when using the Ochrophyta primer suite.

Phylogenetic and genetic variation analyses

GenBank entries are available for specimens expertly identified as *Corallina* (prior to this study). Thus, it was expected that there would be representative short-read sequences (i.e. ASVs) generated in this study, likely to have a great number of reads (at least for COI and Rhodophyta primers), taxonomically assigned as *Corallina* with a very high or exact identity match. These sequences (i.e. one ASV per primer set per specimen) were identified and exported for all primer sets.

For *rbc*L, the short-read and the Sanger sequences were aligned per specimen (standard Geneious alignment; default parameters). The consensus sequences were then used in a multiple alignment with *Corallina* reference sequences pulled from GenBank (species-level) that had been generated in the most recent *Corallina* phylogenetic publications, which included reference sequences that originated mostly from expertly identified voucher material and were rooted by type specimen sequence data (Calderon et al. 2021; Schipper et al. 2023; Wade et al. 2023; Nelson et al. 2025; Wade et al. 2025; Suppl. material 1: table S2). To retain as many GenBank-pulled reference sequences as possible, the consensus sequences were slightly trimmed at the 5’ end (28 bases) before exporting the sequence alignment. Further to this, copies of the full *Corallina* alignment were trimmed to represent amplified portions of short-read primers (i.e. Rho1-3, Rho5-7, Rho7-8) and exported. Where possible, *Corallina* species names were updated according to most recent literature and any misspellings corrected and the aligned fasta file exported for analyses. Sequences were imported into R and checks were performed to ensure that all sequences were assigned at species-level (removing *Corallina* sp. and ambiguous clade assignments), had the same length and did not contain any ambiguous bases.

To evaluate variation within primer sets and, therefore, to be able to assess their utility for species level identification, single nucleotide polymorphisms (SNPs) were investigated. SNPs were extracted from the sequences (*fasta2DNAbin*; adgenet v.2.1.11) and scaled allele frequencies were computed

(*scaleGen*). A principal component analysis (PCA) was then performed (*dudi.pca*; ade4 v.1.7-23) and visualised with ggplot2. The plots were also viewed in 3D to explore spatial separation not easily captured in 2D plots (*plot_ly*; plotly v.4.11.0). Any species, and/or individuals within a species, with the same sequences were grouped together so that a single representative sequence was plotted per group. Noting the separation visible in 3D that was not easily comprehended with multiple 2D PCA plots, the first three principal components were used to calculate the minimum distance between species (*dist*; method = Euclidean; stats v.4.4.1). For every species, the five nearest neighbouring species were identified for plotting in ggplot2 (nearest-neighbour plot). The median distance to the closest neighbour was also calculated.

RAxML (v.8.2.11; Stamatakis (2014)) via the Geneious plugin, was used to build the Maximum Likelihood phylogenetic trees (Rho1-8 for *rbc*L). Trees were built using the nucleotide model GTR GAMMA I (see Tree_model_comp.R script for more details), using default settings, but increasing the bootstrap value to 1,000 (set seed 369). *Crusticorallina muricata* (Foslie) P.W.Gabrielson, Martone, K.R.Hind & C.P.Jensen and *Ellisolandia elongata* (J.Ellis & Solander) K.R.Hind and G.W. Saunders were included, specifying *E. elongata* as the outgroup (Schipper et al. 2023). The trees were exported in newick format and imported into R to create figures using ggtree (v.3.17.0; Yu et al. (2017)).

The same process and analyses were done for COI (Table 1), but differed by using the most abundant COI representative sequence for each, as no ASVs for Carm specimens were assigned to *Corallina*. To be able to retain as many aligned GenBank *Corallina* sequences as possible, four bases were trimmed from the 3' end. In addition, for voucher specimens that had GenBank entries for both COI and *rbc*L (Rho1-8), the sequences were concatenated and the SNPs extracted to then generate a PCA and nearest-neighbour plot.

While overall patterns of species placement were visualised through PCAs, statistical testing was approached with caution due to the relatively low replicate numbers for some species. To assess significant genetic divergence amongst species, permutational multivariate analysis of variances (PERMANOVA) was conducted. Pairwise genetic distances were calculated using both COI and Rho1-8 SNP data (*dist.dna*, Kimura 2-parameter model; ape v.5.8-1). These distance matrices were analysed using *adonis2* (vegan v.2.6-10), followed by pairwise comparisons via *pairwise.adonis* (pairwiseAdonis v.0.4.1). Multiple comparisons were corrected using the false discovery rate method. For pairwise PERMANOVA, only species with more than four replicates were included to increase output reliability. To capitalise further on the produced genetic distance datasets, species delimitations were visualised to look at genetic proximity amongst *Corallina* species sequences clustering *de novo*. Briefly, the genetic distance distribution was calculated and assessed for the difference between the intraspecific and interspecific divergence. Hierarchial clustering was performed using an unweighted pair group method with the arithmetic mean method (*hclust*, method = average; stats) and differences in clustering investigated (*cutree*, heights parameter used; stats). An appropriate clustering threshold was set after differences were assessed and clustering displayed using a heatmap (ggplot2). These results were then compared to literature that had used more traditional species delimitation methods.

Primer design, pairing and testing

Chlorophyta, Ochrophyta and Rhodophyta gene region selection

Expanding on van der Reis et al. (2023), a comprehensive list of known New Zealand seaweed species (Nelson et al. 2023; species of Phaeophyceae were the predominant Ochrophyta targeted) was run through the GenBank nucleotide database (not limited to New Zealand being the geographic location; July 2024) using Entrez Direct (v.13.3; Kans (2013)). Species entries were pulled from GenBank, if possible, for 16S rRNA, 18S rRNA, 23S rRNA, COI, ITS, *psb*A, *rbc*L, *trn*L and *tuf*A entries (Fig. 2a; van der Reis et al. (2023)). The available sequences were downloaded with selected associated metadata (e.g. accession identification, full taxonomy, publication status, geographic location), from which a fasta file was also created. The number of entries per species was calculated and *rbc*L was notably abundant in species entries across Chlorophyta, Ochrophyta and Rhodophyta. Thus, *rbc*L was the gene region selected for exploratory modular universal primer design for future barcoding and metabarcoding. To note, literature was initially searched for appropriate universal *rbc*L primer sets, but the majority developed were either for barcoding purposes targeting entire gene regions (produced amplicons longer than we desired; e.g. Mshiywa et al. (2024)), were not developed for macroalgae (e.g. Hadi et al. (2016)) or had non-specific amplification (e.g. Akita et al. (2020)). We soon realised that, to have a modular universal primer suite, it was going to be more fruitful to design our own.

*rbc*L primer design and selection

For every species in the *rbc*L fasta file, up to five sequences were randomly selected. Subsequently, the file was split by species into Chlorophyta, Ochrophyta and Rhodophyta for ease of primer development. MAFFT (v.7.505; Katoh and Misawa (2002)) was used to align the sequences, implementing the 'adjust direction' parameter. The alignments were then passed into Geneious. A consensus sequence was made using a 90% threshold, ignoring gaps (and hiding sites with over 90% gaps) and suitable regions for primers identified which were ideally capable of pairing with a neighbouring primer to ensure a target region size suitable for sequencing on an Illumina platform (2 × 300 bp) with minimal ambiguous bases.

All *rbc*L GenBank entries for Chlorophyta, Rhodophyta and Ochrophyta were retrieved (April 2025) with their metadata (Fig. 2b). Thereafter, a *rbc*L fasta database was created and assessed for entries that could be captured by the designed primers – tested *in silico* using Cutadapt with the parameters no indels or errors allowed. The primers were also assessed against standard primer guidelines (e.g. Robertson and Walsh-Weller (1998); Prediger (2013); Basu (2015)). For example, the melting temperatures of possible hairpin and self-dimer formations needed to be low such that the primer annealing temperature would remain above them and the GC content ranging between 40–60%. Similarly, primer pairs were run through the *rbc*L databases and only the sequences for which both the forward and reverse primers matched were retained (using Cutadapt). Three primer pairs that spanned across the *rbc*L gene region for each phylum were

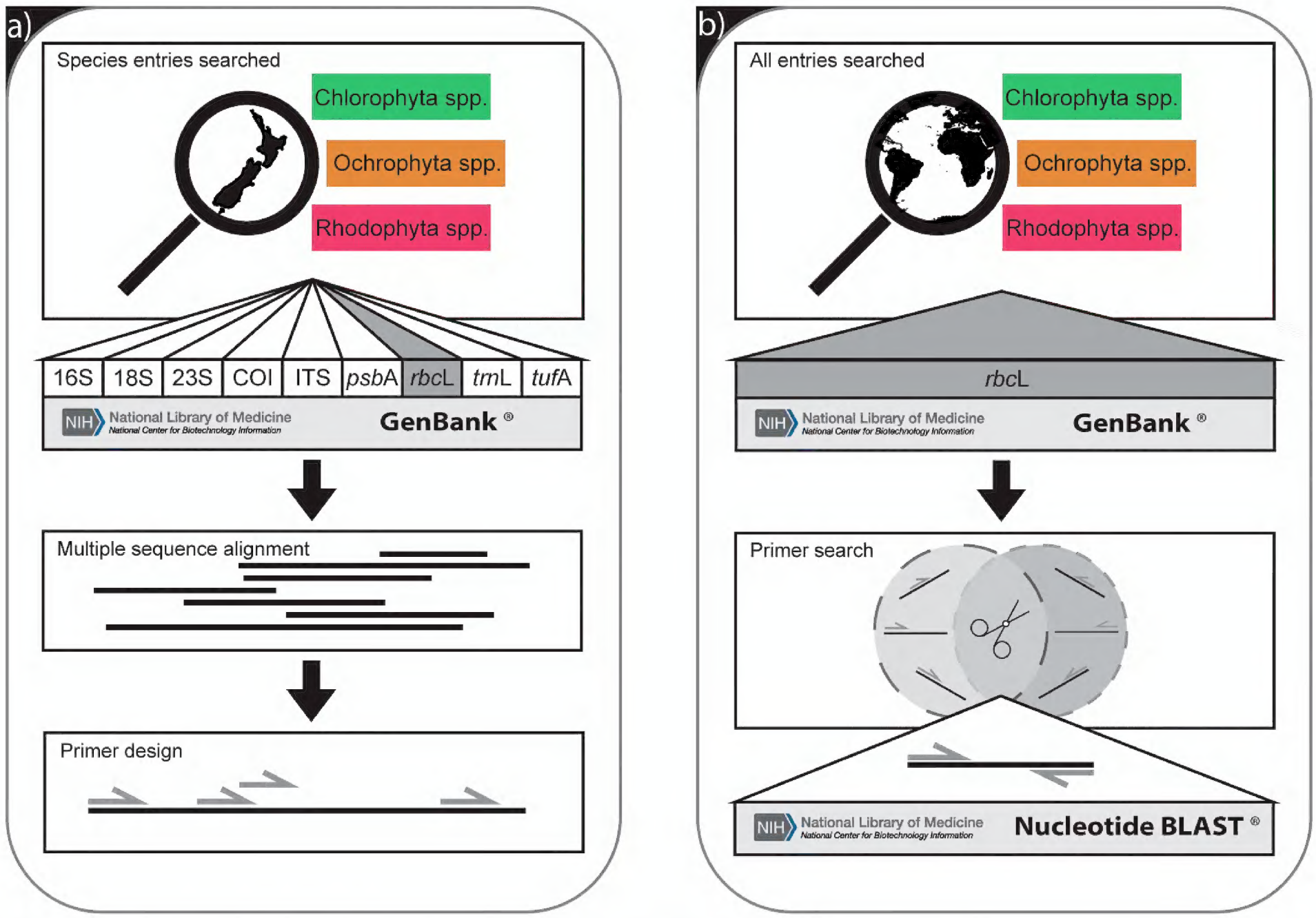

Figure 2. The primer design, development and testing workflow. **a**) New Zealand macroalgae species were searched in the National Center of Biotechnology Information's nucleotide database (GenBank) for nine gene regions. It was confirmed that *rbc*L had the greatest number of species entered, the sequences and selected metadata were pulled for primer design and development. The multiple alignment, per phylum, provided a consensus sequence that was used for primer design; **b**) GenBank *rbc*L sequence entries and selected metadata were pulled from GenBank for Chlorophyta, Ochrophyta and Rhodophyta. The designed primers were searched against the custom database for matches. Primer sets were also run against the databases, trimming to keep the fragment between the primers. These sequences were then assigned taxonomy using MegaBLAST to assess possible taxonomic assignment levels.

selected (i.e. Chlorophyta, Ochrophyta and Rhodophyta). These primer pair fasta files were then re-assigned taxonomy using MegaBLAST to assess their specificity regarding taxonomic assignment levels (i.e. species, genus or higher).

Results

Corallina armata case study

The three Rhodophyta *rbc*L primer sets amplified more readily on the *Corallina* specimens (Carm1-4, Cber1) than the Ochrophyta *rbc*L primers sets (single bands indicating primer set specificity as expected). The total number of raw reads was 83,775 for COI, 100,595 for 'Rho' and 37,706 for 'Och' (Tables 1, 2). Rho5-7 and Och6-9 had the greatest number of raw reads across all samples, whereas Rho1-3 and Och1-3 had the least for Rhodophyta and Ochrophyta, respectively (Table 2; Suppl. material 1: table S3). There were no reads remaining for any of the controls post primer removal, except for CONTROL_Och6-9;

Table 2. Processing of sequencing data by primer set from raw to final dataset. The raw reads were run through Cutadapt (v.4.1; Martin (2011)), searching for primer pairs (Rho and Och primers were pooled independently prior to sequencing). The sequences were retained for further processing if both the forward and reverse primers matched. These sequences were then used as the input data for DADA2 (within Qiime 2 v.2023.5 – Bolyen et al. (2019); range and median values provided). The reads and the total number of amplicon sequence variants (ASVs; in brackets) are provided for sequences passing DADA2's sequencing quality filtering and the minimum reads and relative read abundance thresholds in R.

Primer set	Raw reads	DADA2 input	Range	Median	Post-DADA2	Post-R
COI (Leray)	83,775	69,364	9,229–17,625	15,314	66,594 (53)	65,503 (14)
Rho1-3	100,595	3,505	341–1,520	555	3,273 (18)	3,261 (15)
Rho5-7		46,880	8,088–10,805	9,485	44,885 (26)	44,506 (8)
Rho7-8		27,764	4,160–6,581	5,533	26,908 (12)	26,876 (10)
Och1-3	37,706	643	59–315	71	437 (18)	399 (11)
Och3-6		2,759	90–1,394	249	2,567 (27)	2,549 (23)
Och6-9		24,902	1,290–12,938	3,909	19,940 (52)	19,694 (37)

however, all reads were filtered out upon quality filtering in DADA2 (Suppl. material 1: table S3). Post-DADA2, an average of 95% of reads were retained after filtering for the *rbc*L Rhodophyta primer sets, 73% for the *rbc*L Ochrophyta primer sets and 96% for COI. The targeted *rbc*L fragments matched their expected bp sizes (Table 1), but there was variation amongst summary statistics (mean, minimum and maximum; median provided the correct expected size) when no filtering of minimum reads and RRA per ASV per sample occurred. The filtering reduced the number of reads by less than 2% for each primer set, except for Och1-3 which was reduced by 8%. On the other hand, the number of ASVs that remained varied by primer set. COI and Rho5-7 were reduced by approximately 70%, Och1-3 by 39%, Och6-9 by 29% and the others by about 15% (Table 2).

For *rbc*L, one ASV for each Rhodophyta primer set was assigned to *Corallina*, whereas only some Ochrophyta primer sets had ASVs assigned to *Corallina* (consistency varied amongst samples; Fig. 3). Carm1–4 matched a single GenBank entry collected from New Zealand with 100% confidence (GenBank accession: KM369146; Rho7-8 results differed due to resolution). The GenBank entry had been misidentified as *C. cæspitosa* R.H.Walker, J.Brodie & L.M.Irvine. The type material of *C. cæspitosa* has been shown to be synonymous with the earlier named *C. berteroi* and differs from the sequence data for Carm1-4. The *Corallina* sp. assignment for Cber1 was composed of multiple *Corallina* GenBank entries at 100% confidence (e.g. *Corallina ferreyræ* E.Y.Dawson, Acleto & Foldvik, *C. cæspitosa*, *Corallina* sp. Clades, *C. berteroi*). *Corallina ferreyræ* is also regarded as a synonym of *C. berteroi*. For ease of differentiation throughout the manuscript, "*C. cæspitosa*" is used when referring to GenBank matches where the name has been incorrectly applied and/or specimens fall outside the *C. berteroi* clade (refer to section '*Corallina* conundrums' regarding incorrectly applied names).

The ASVs with the greatest RRA percentages for the Rhodophyta primer sets were assigned as *Corallina*. One exception was Carm2 for Rho1-3. For the Ochrophyta primer sets, the RRA varied for ASVs assigned as *Corallina*. Regardless, Och6-9 was the Ochrophyta primer set that provided the most consistency in producing a *Corallina* assigned ASV amongst samples. Carm2 was

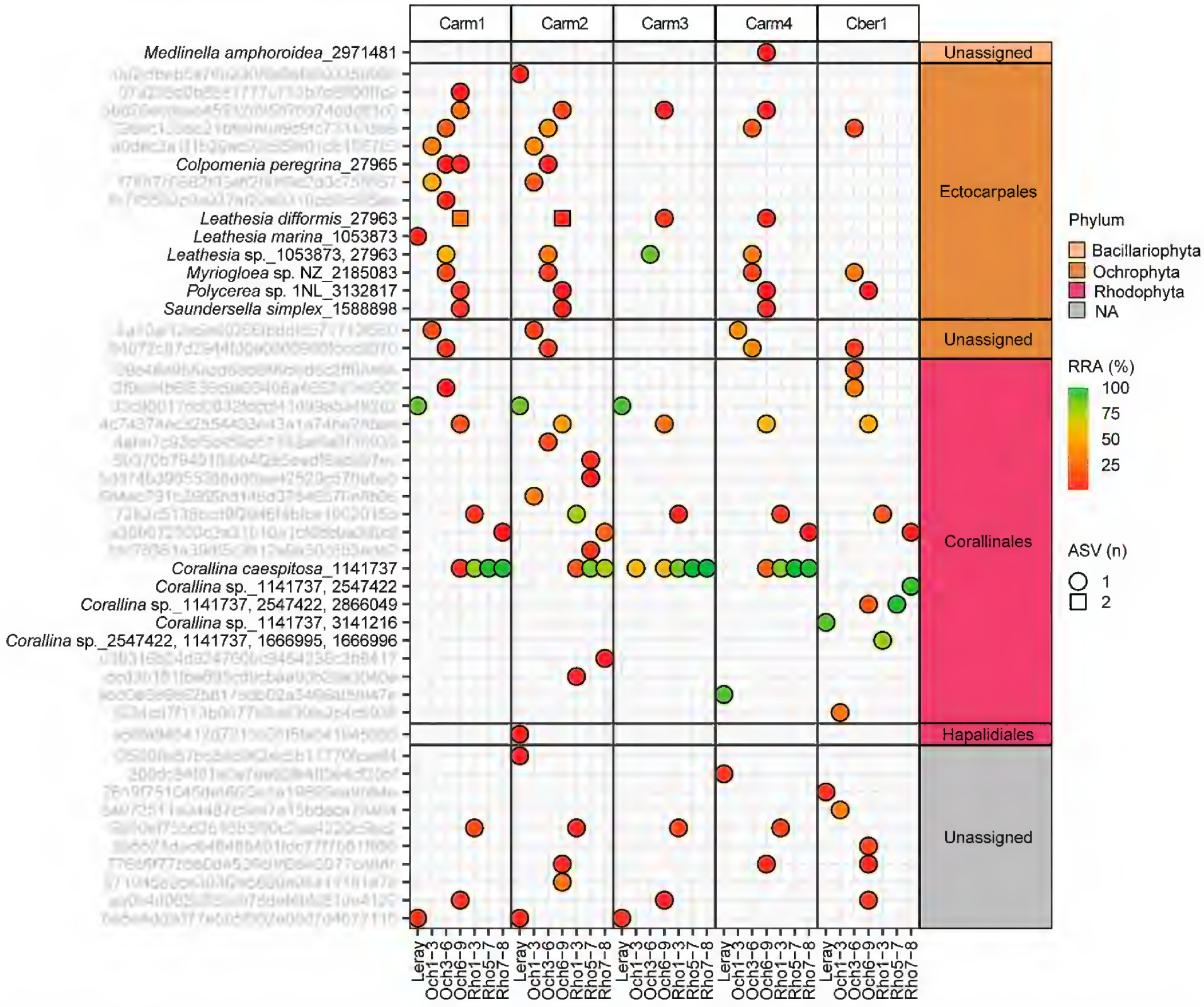

Figure 3. ASVs detected per *Corallina* sample (*C. berteroi*: Cber1; *C. armata*: Carm1–4). Two genes were targeted, COI and *rbc*L, using universal primers (x-axis). Modular primer suites were designed for Rhodophyta and Ochrophyta, which spanned the gene region. ASVs that had taxonomy assigned with a minimum percentage of 97% are labelled in black and the others are in grey (left y-axis). ASVs were grouped if assigned to the same species or when a species could not be determined, but assigned to the same genus (single ASV represented by a circle and two by a square). ASVs have been grouped by phyla they were assigned to, if possible, these are colour coded and the order is provided (right y-axis). The relative read abundance (RRA) percentage is indicated by the shape colour.

the only sample where no ASV could be assigned for Och6-9. Och1-3 also provided a *Corallina* assigned ASV for Carm3. The filter parameters indicated that filtering to retain a minimum of 10 reads removed the ASV from Carm4 and Cber1. Investigating primer matches to the consensus sequence revealed that the Rhodophyta primer suite had no mismatches, whereas for the Ochrophyta primer suite Och1 and Och9 were the only primers that matched exactly.

The only COI ASV assigned as *Corallina* was for Cber1 (for multiple GenBank entries as above), but for all samples, there was a marked difference between the ASV with the greatest RRA percentage and all the other ASVs (Fig. 3; Suppl. material 1: table S4). The taxonomic assignments for the ASVs with the greatest RRAs for Carm1–4 did not pass the confidence threshold for species assignment (*Corallina chamberlainiæ*, 95%; now *C. hombronii*, Nelson et al. (2025)). The *Corallina* ASV for Carm4 differed at one base (position 286 of the 313 bases) compared to Carm1–3.

Sanger sequencing resulted in high quality sequences, except for Carm2 (despite the repeated sequencing). All short-read ASVs assigned as, or assumed to be (high RRA percentage), *Corallina* were aligned with their respective Sanger sequences. Carm2 consensus sequence was the only one that needed inspection as there were 'base disagreements' between the short-reads and the Sanger sequences. The short-reads were primarily relied upon for corrections in the initial alignment, other problematic base calls were inspected by comparing Carm1's consensus sequence and were corrected (three bases). The final consensus sequences aligned and there were no disagreements for the *C. armata* samples, but there were 12 bases that differed when compared to *C. berteroi* (1% of the amplified region; Suppl. material 1: table S5). Between these two *Corallina* species, Rho1-3 and Rho7-8 differed at two bases (0.6% and 0.7%, respectively) and Rho5-7 differed at six (2%). In comparison, COI had 18 base differences between the two species (5.8%).

The final sequence length used for the COI analyses was 309 bases (217 sequences representing 17 species, 85 SNPs; Suppl. material 1: table S6). For the *rbc*L analyses, the final length was 297 bases for Rho1-3 (103 sequences representing 16 species, 34 SNPs), 299 bases for Rho5-7 (119 sequences representing 16 species, 46 SNPs) and 273 bases for Rho7-8 (226 sequences representing 17 species, 29 SNPs) (Suppl. material 1: table S6). Rho7-8 was the only *rbc*L primer set to include *C. saundersii* R.M.Wade, K.R.Hind, Ogushi & Martone, due to the region targeted (3' end) by the primer set. For the longer Rho1-8, the total length used was 1,130 bases (86 sequences representing 16 species, 126 SNPs). The total number of SNPs per primer set is likely inflated due to incorrect base calls in reference sequences (see section '*Corallina* conundrums').

Polytomy occurred for both COI and *rbc*L trees (Figs 4, 5), although more substantial in COI, poorly resolving deeper nodes. Regardless of the polytomy, species consistently clustered together in species groups for both gene regions. The material collected as *C. berteroi* (Cber1) in this study was part of the *C. berteroi* clade, which included *C. yendoi* Martone, P.W.Gabrielson, M.S.Calderon & D.E.Bustamante for both gene regions. The material collected as *C. armata* (Carm1–4) from Mataikona in this study was placed in a clade with "*C. cæspitosa*" for *rbc*L. No sequence entries were available for "*C. cæspitosa*" for COI and *C. armata* was identified as a sibling clade to *C. hombronii*. *Corallina hombronii* was a sister clade to *C. berteroi* for *rbc*L.

The PCAs typically mirrored similar species clustering as the trees, but differed in spatial placement between *rbc*L and COI regions and within *rbc*L given multiple primers were used targeting different regions (Figs 4, 5). The 3D visualisations and the ability to interact with the plots made them very informative for a realistic spatial plotting of points. The 3D spatial separation amongst species was effectively transformed into 2D (Figs 4c, 5c). *Corallina bathybentha* E.Y.Dawson, *C. chilensis* Decaisne, *C. maxima* (Yendo) K.R.Hind & G.W.Saunders and *C. armata* had the greatest separation from other species for COI (Fig. 4b, c). The species for *rbc*L (Rho1-8) with the greatest separation were *C. officinalis* and *C. americana* R.M.Wade, Martone & P.W.Gabrielson, *C. pilulifera* Postels & Ruprecht and *C. hakodatensis* Martone & R.M.Wade. Using combined SNPs of COI and *rbc*L (same voucher specimens; 32 sequences representing 12 species, 163 SNPs), five distinct groups were identified (Fig. 6). *Corallina armata*, *C. berteroi*, *C. hombronii* and *C. yendoi* formed one group, with species clustering separately; similarly seen in

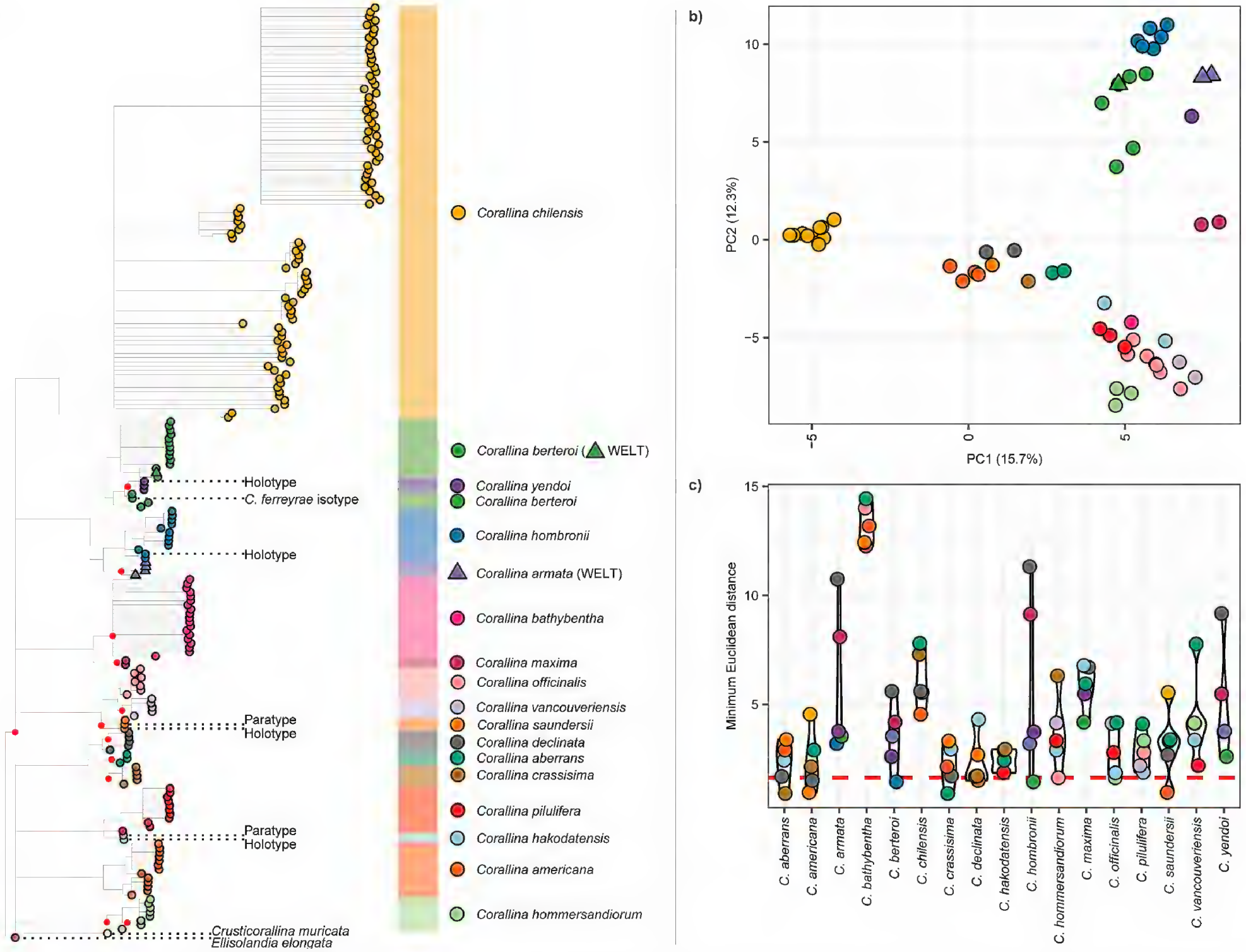

Figure 4. COI phylogenetic tree (left; **a**) PCA plot (top right; **b**) and nearest-neighbour PCA-derived plot (bottom right; **c**) for *Corallina* voucher specimens. **a**) The phylogenetic tree used a 309 base portion of the COI gene region and was built using RAxML. Bootstrap values that had a minimum of 90% are indicated on the tree as a red dot. Voucher specimen sequences that are from type material are indicated as such. The *Corallina* species colours used in the PCA match the species colours used in the tree. The voucher specimens for *C. armata* and *C. berteroi* (WELT) used in this study are represented as triangles and all other specimens as circles (except nearest-neighbour plot); **b**) The PCA plot indicates the spatial arrangement of the *Corallina* species using principal coordinates one and two (and the percentage variation that can be explained by each); **c**) The nearest-neighbour plot is a PCA-derived plot that shows which species are situated closest to one another in a 3D space. The five closest neighbours are plotted for each species and the red dashed line represents the closest neighbour median.

Metabarcoding and Metagenomics 10: 69–104 (2026), DOI: 10.3897/mbmg.10.173634

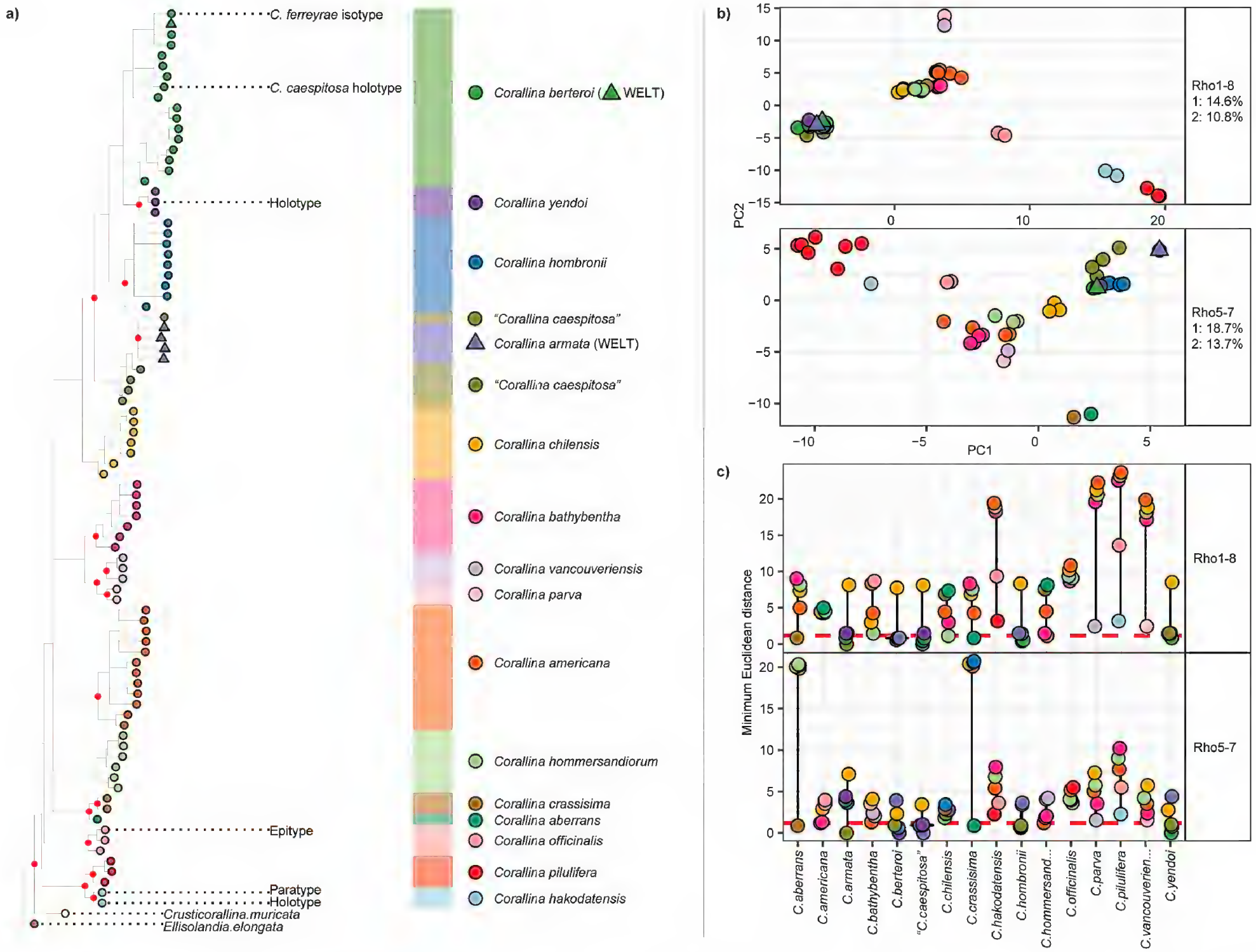

Figure 5. *rbc*L phylogenetic tree (left; **a**) PCA plot (top right; **b**) and nearest-neighbour PCA-derived plot (bottom right; **c**) for *Corallina* voucher specimens. **a**) The phylogenetic tree used a 1,130 base portion of the *rbc*L gene region and was built using RAxML. Bootstrap values that had a minimum of 90% are indicated on the tree as a red dot. Voucher specimen sequences that are from type material are indicated as such. The *Corallina* species colours used in the PCA match the species colours used in the tree. The voucher specimens for *C. armata* and *C. berteroi* (WELT) used in this study are represented as triangles and all other specimens as circles (except nearest-neighbour plot); **b**) The PCA plot indicates the spatial arrangement of the *Corallina* species using principal coordinates one and two (and the percentage variation that can be explained by each); **c**) The nearest-neighbour plot is a PCA-derived plot that shows which species are situated closest to one another in a 3D space. The five closest neighbours are plotted for each species and the red dashed line represents the closest neighbour median.

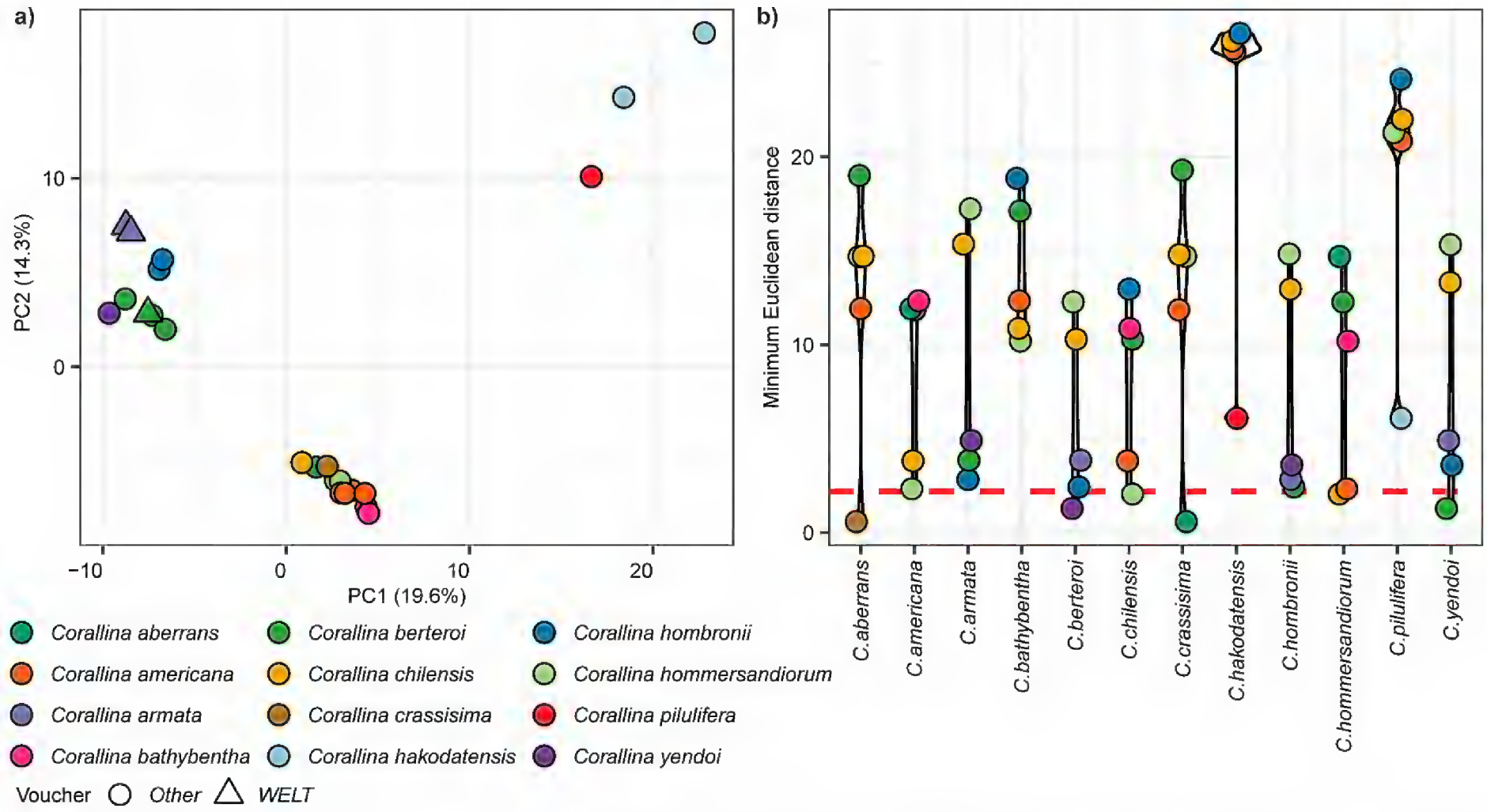

Figure 6. PCA plot (left; **a**) and nearest-neighbour PCA-derived plot (right; **b**) for *Corallina* voucher specimens that had sequences available for both *rbc*L (Rho1-8) and COI. The voucher specimens for *C. armata* and *C. berteroi* (WELT) used in this study are represented as triangles and all other specimens as circles in the PCA plot. For the nearest-neighbour plot, only circles are used. The red dashed line represents the closest neighbour median.

the independent COI and *rbc*L plots. A second group is formed by *C. bathybentha*, largely separated from other species in COI rather than *rbc*L. The remaining three groups are defined more so with *rbc*L than COI. *Corallina pilulifera* and *C. hakodatensis* group together, although there is clear species separation within the group. Another group is composed of *C. chilensis*, *C. americana* and *C. hommersandiorum* R.M.Wade, Martone & P.W.Gabrielson and the final group by *C. aberrans* (Yendo) K.Hind and G.W.Saunders and *C. crassisma* (Yendo) K.Hind and G.W.Saunders. For the short-read Rhodophyta primers, this clustering is most like Rho5-7 (Fig. 5).

The PERMANOVA indicated that there was overall significant genetic distance amongst species (*adonis2*; *rbc*L (Rho1-8): $F = 240.39$, $R^2 = 0.98$, $p < 0.001$; COI: $F = 1240.6$, $R^2 = 0.99$, $p < 0.001$) and further pairwise testing indicated all species (species with $n \geq 4$) comparisons had significant separation, based on genetic distance, except *C. armata* versus "*C. cæspitosa*" for Rho1-8 (Suppl. material 1: table S7; see section '*Corallina* conundrums'). Furthermore, sequences clustered together in species groups *de novo*, providing confidence in species delimitations (as in previous *Corallina* studies; e.g. Calderon et al. (2021); Wade et al. (2023); Wade et al. (2025)) and the confirmation of the identity of *C. armata* (Suppl. material 1: figs S1, S2).

Overall, Rho5-7 was the most informative short-read Rhodophyta primer. Sequence exact matches between species occurred for individuals from *C. berteroi* and *C. yendoi* and "*C. cæspitosa*" and *C. armata*. Rho1-3 was the second most informative short-read species marker (exact matches: *C. berteroi* and *C. hombronii*, "*C. cæspitosa*" and *C. berteroi* and "*C. cæspitosa*" and *C. armata*). Rho7-8 was the least informative, although still ideal for some species identification (Suppl. material 1: fig. S3). In general, COI proved to be more powerful for species clustering given the greater genetic distances identified using

this gene region in comparison to *rbc*L, which had less well-defined intra- and inter-specific divergence (Suppl. material 1: figs S1, S2).

The high sensitivity of next-generation sequencing, combined with the universal primer design, enabled the detection of additional algal taxa likely associated with the specimens or their surrounding environment. In general, Ochrophyta primers amplified Ochrophyta taxa and Rhodophyta primers amplified Rhodophyta taxa (Suppl. material 1: table S4). Overall, there were few species-level assignments to sequences at the required confidence threshold, but the majority of the community assignments belonged to Ochrophyta. *Medlinella amphoroidea* Frankovich, Ashworth & M.J. Sullivan, *Polycerea* sp. and *Saundersella simplex* (De A.Saunders) Kylin, species which are not known from New Zealand, but matched at 99–100% (except *M. amphoroidea* at 98%).

Corallina armata epitypification

The results from this study provide evidence of a distinctive species of *Corallina*, conforming to the description of *C. armata* from the southern east coast of North Island of New Zealand, given the position of specimen sequences in unique clades in the phylogenetic trees and the overall separation from other species when comparing genetic distance and the PCA/nearest-neighbour PCA-derived plots (Figs 4–6 and Suppl. material 1: figs S1, S2). For this reason, an epitype specimen has been selected for this species, following the requirements of the Code (Art. 7.11; Art. 9.9; Turland et al. (2025)), to serve as "an interpretative type when all original material associated with a validly published name, is demonstrably ambiguous and cannot be critically identified for purposes of the precise application of the name to a taxon".

***Corallina armata* Hook. f. & Harv., Harvey 1849: 103, Tab. XL**
Figs 1–7

Holotype. Not designated; Original material: Colenso No. 224 and 654 in TCD and BM.

Lectotype (designated here): Colenso specimen 654, TCD0011847, (annotated by H.W. Johansen, 22 Sep 1967).

Epitype (designated here): WELT A036058 (Suppl. material 1: fig. S4), New Zealand, North Island, Wairarapa, Mataikona, (-40.803 S, 176.260 E), 08 February 2024, leg. W.A.Nelson, AWA001. Mid-low intertidal to upper subtidal, on reef platform, in channels and pools.

Primer design, pairing and testing

The number of GenBank species (n = 808; New Zealand species list; Nelson et al. (2023)) that had entries were calculated for the nine selected gene regions (Suppl. material 1: table S8). The top three gene regions overall were *rbc*L (n = 438 species), COI (n = 229) and *psb*A (n = 209) (Fig. 7). *rbc*L also had the greatest number of species that had entries for Chlorophyta (n = 57 species), Ochrophyta (n = 100) and Rhodophyta (n = 281) when separated by phylum (Fig. 7 and Suppl. material 1: table S8). This was also reflected when calculating entries whose geographic location had been entered as New Zealand.

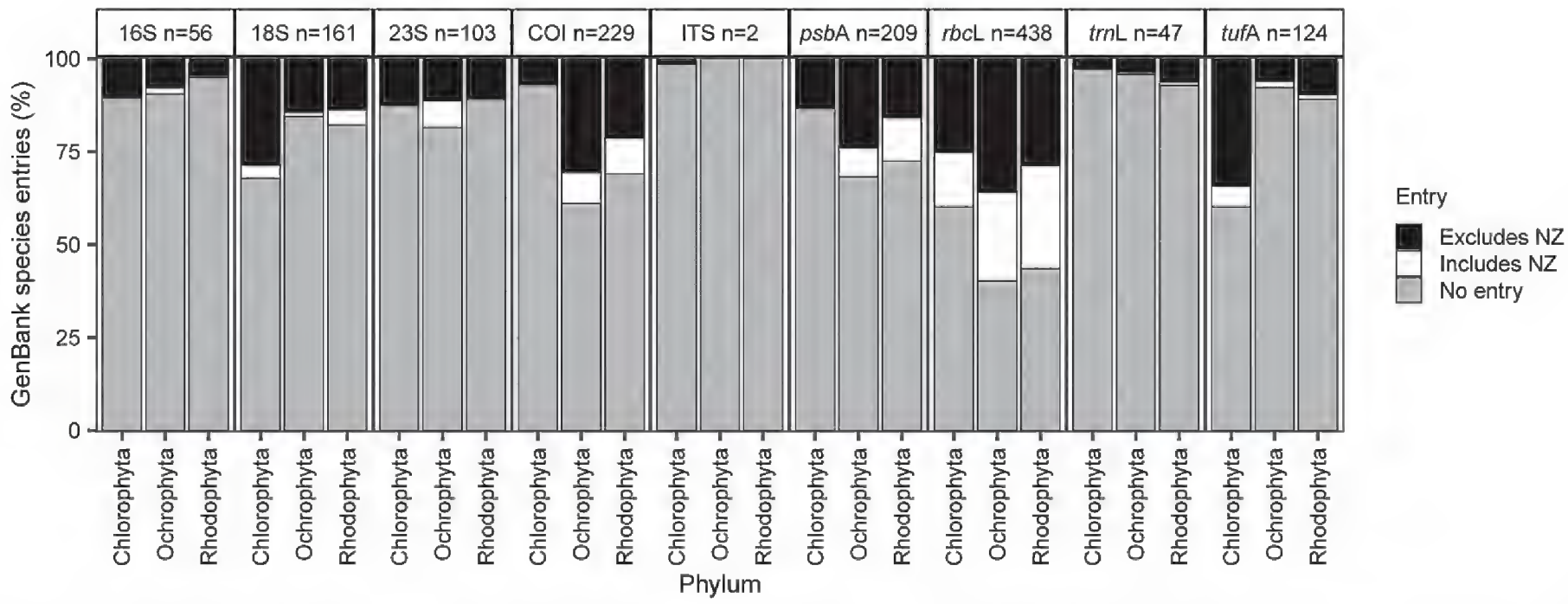

Figure 7. GenBank entries for New Zealand (NZ) listed species (Nelson et al. 2023) available for nine gene regions. The number of species with entries is provided next to the gene name. The species have been categorised as either having entries, but geographic location not listed as New Zealand (i.e. 'Excludes NZ'), having entries with at least one entry that lists New Zealand as the geographic location (i.e. 'Includes NZ') or no entry for the species (i.e. 'No entry').

In total, there were 11 primers designed for Chlorophyta, 12 for Ochrophyta and nine for Rhodophyta, which spanned over 1,000 bases of the *rbc*L gene region (Fig. 8 and Suppl. material 1: table S9). Regardless of the primers *in vitro* suitability statistics (e.g. hairpin and self-dimer melting temperatures), it was found the primers had the potential to match a large number of GenBank *rbc*L sequence entries for the three phyla (Suppl. material 1: table S10, figs S5–S7). On average, the Rhodophyta primers captured approximately 60% of Rhodophyta GenBank entries each. Chlorophyta and Ochrophyta primers captured less at 40% and about 30% each for the Chlorophyta and Ochrophyta GenBank entries, respectively. Overall, about 1% of phyla entries were not captured by any of their respective primers (Table 3). The potential cross-phyla matches indicated that Chlorophyta primers were the most phylum-specific (Chl4 and Rho5 have similar sequences; Fig. 9). Ochrophyta and Rhodophyta primers tended to be less specific, with Rhodophyta primers matching the Ochrophyta entries more readily, although greater specificity was found for the New Zealand database

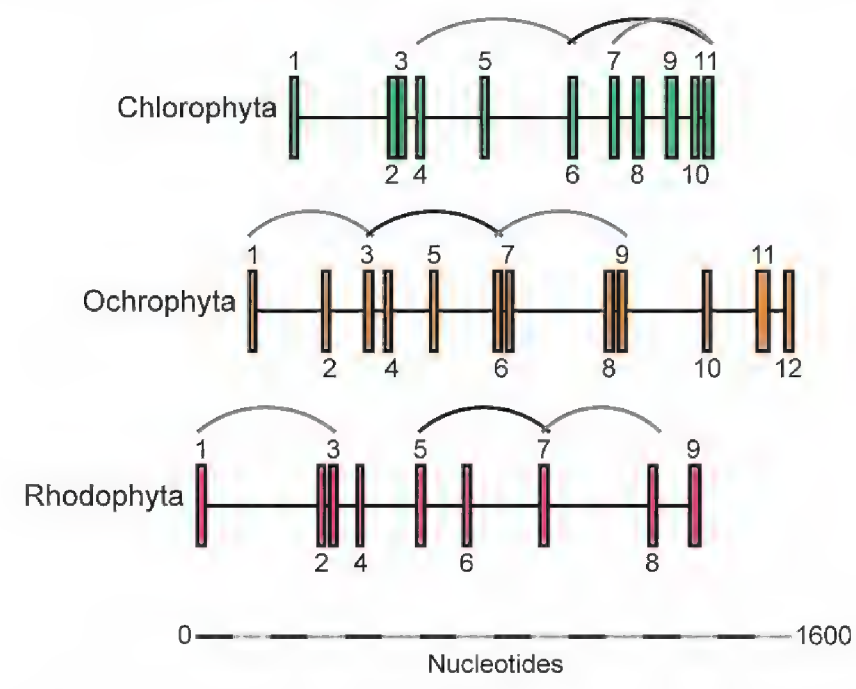

Figure 8. Primer positions along the *rbc*L gene region. The overall alignment of Chlorophyta, Ochrophyta and Rhodophyta primer suites were based on primer similarity (i.e. Chl4 and Rho5) and matches amongst phyla (i.e. Ochrophyta and Rhodophyta). The arcs join the primer pairs used in this study.

Table 3. GenBank entries available (April 2025) for Chlorophyta, Ochrophyta and Rhodophyta and how well the respective *rbc*L primers designed in this study could capture them. The primer match ranges, mean and no match statistics are provided.

Primer target	GenBank entries	Match range	Match mean	No match
Chlorophyta	12,737	1,179–8,746	5,155	104
Ochrophyta	16,033	1,713–10,581	5,332	136
Rhodophyta	24,902	9,084–18,668	14,678	279

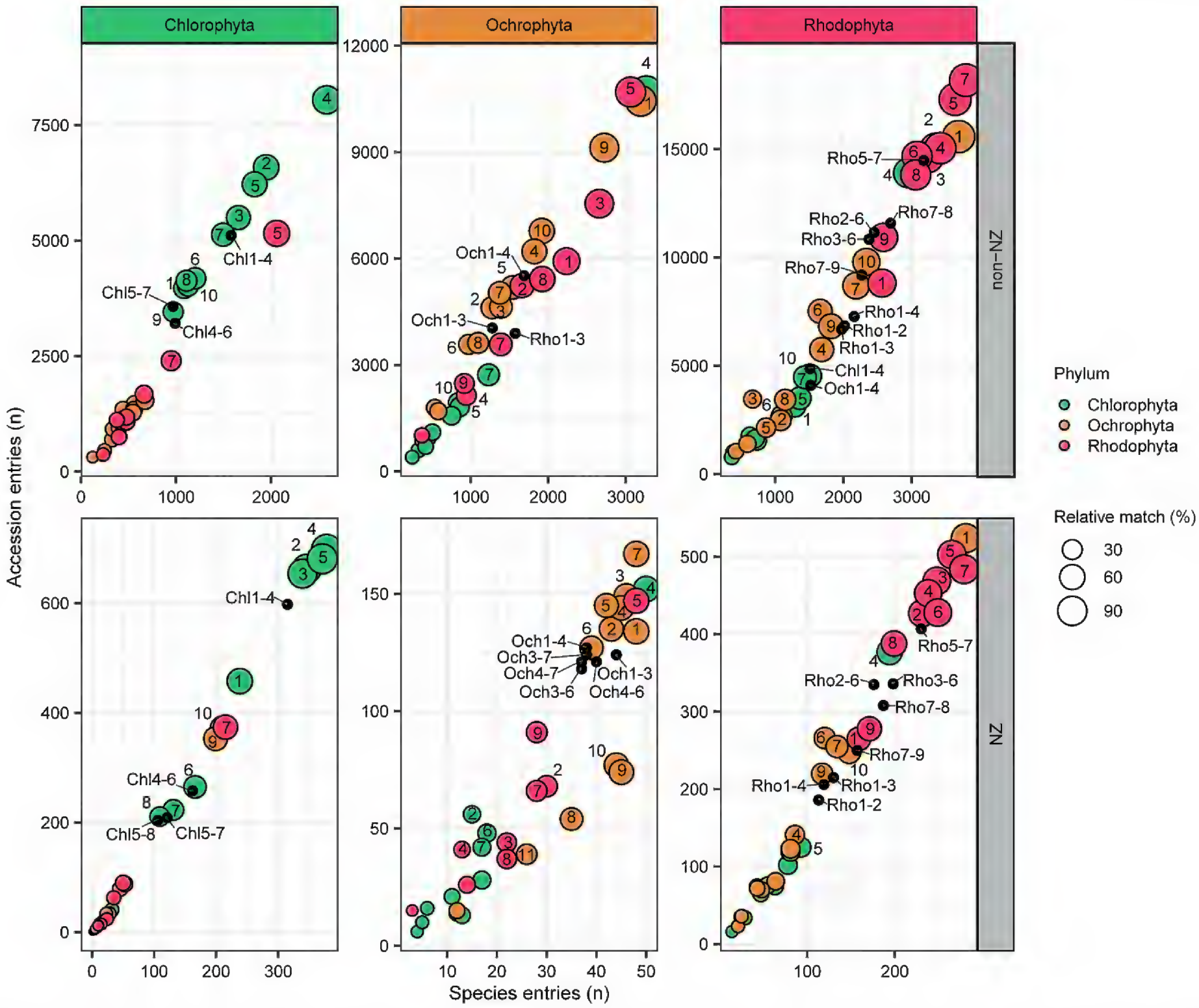

Figure 9. Number of primer and primer pair matches to GenBank and species entries for macroalgae in Chlorophyta, Ochrophyta and Rhodophyta. Results are presented for entries that are listed as New Zealand (NZ) being the geographic location and not (non-NZ). The relative match (represented by circle size) was calculated by taking the total number of species matched per primer divided by the total number of species entries captured (based on sscinames). The fill colour represents algal phyla (i.e. green for Chlorophyta, brown for Ochrophyta, red for Rhodophyta) and the number within/next to is the primer. Primer pairs, that would amplify a region of the gene suitable for short-read sequencing, were assessed for entry matches for all three phyla. For example, for 'Rhodophyta and non-NZ' Rho7 was able to capture the greatest number of entries and species and Rho5-7 was the best performing short-read primer set. Primers and primer pairs are shown if they captured at least 25% of species entries.

entries (Fig. 9). When assessing paired primers targeting short-reads, Chl1-4 captured the greatest number of species for Chlorophyta entries, Och1-4 for Ochrophyta and Rho5-7 for Rhodophyta.

Several primers were 'eliminated' when considering the suitability of each primer pair, for example, approximate annealing temperatures and amplification universality within their respective phyla (Suppl. material 1: table S9). Three primer pairs were selected for each of the phyla with the aim to span the gene region as best possible (Fig. 8, Table 1 and Suppl. material 1: table S9). Capturing the 'amplicons' *in silico* and re-running them through BLAST showed that taxonomic assignment at a genus-level would be most likely for Chlorophyta primer pair amplicons, whereas species-level assignments were more likely for the Ochrophyta and Rhodophyta primer pairs (Fig. 10; Suppl. material 1: table S11). Chl4-6, Och1-3 and Rho5-7 were the primer sets that captured the greatest number of GenBank entries for each of their respective *rbc*L databases.

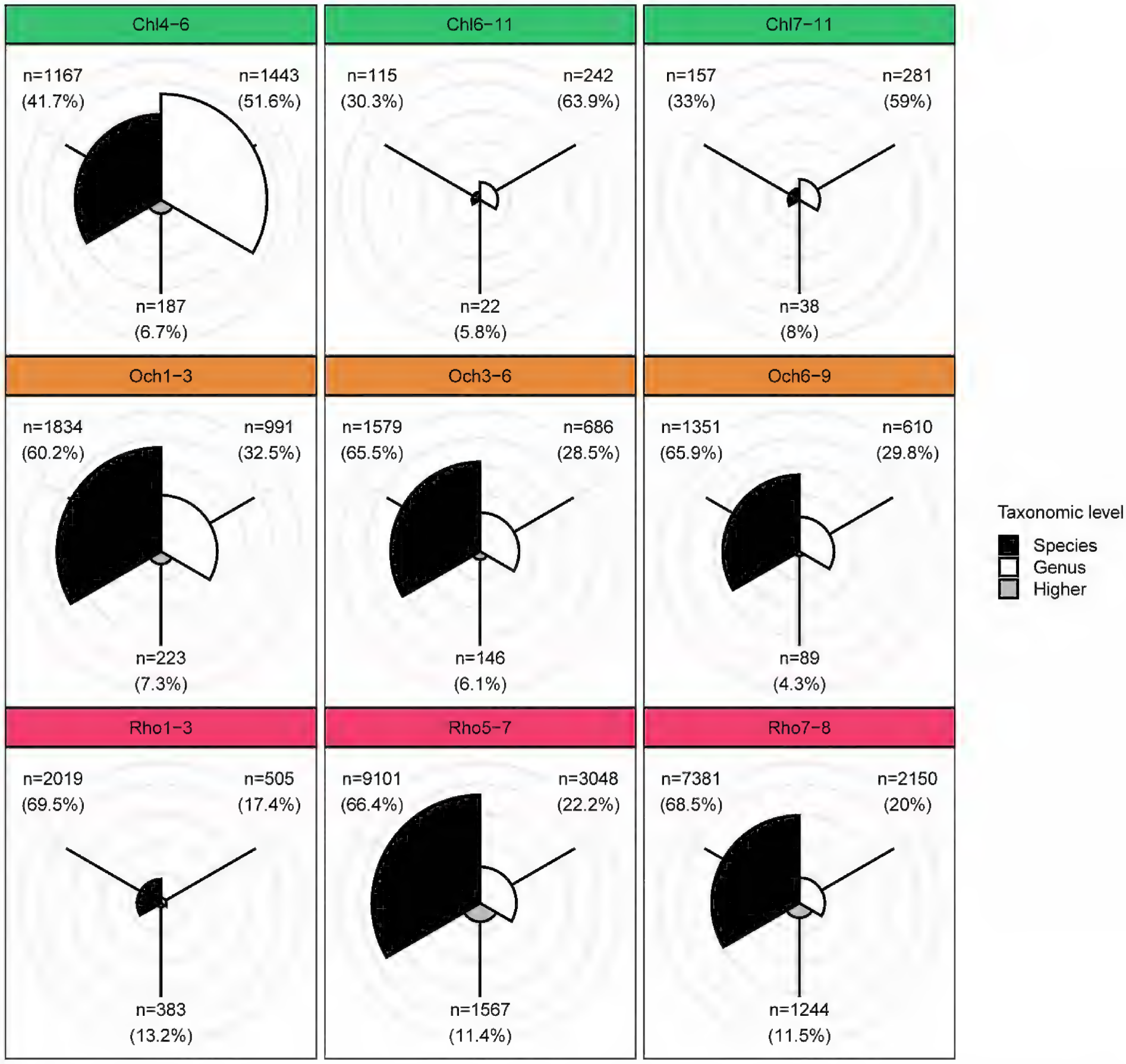

Figure 10. Statistics for the taxonomic assignment levels. Amplicons captured by each primer set were assigned taxonomy using MegaBLAST and compared to their original assignments. The assignments were then categorised as either matching the species they originally were derived from, the genus or a higher level. The top row shows the Chlorophyta primer sets, middle Ochrophyta and bottom Rhodophyta.

Discussion

Here, we provide the epitypification of *Corallina armata*, resolving its taxonomic identity through DNA sequencing of the *rbc*L and COI gene regions for specimens collected from the type locality. By integrating analytical approaches commonly used in population genetics, the research explored species clustering and delineation from each (and both) loci to generate a robust and informative method to leverage SNP data. Additionally, the *in silico* results of modular universal primer suites, initially designed for national application, indicate strong potential for utility in international contexts. The Ochrophyta and Rhodophyta short-read primer sets tested show promise that species-level identification would be possible when applied to other species beyond *Corallina* and they are indicative of being particularly valuable in DNA metabarcoding workflows where sequencing length is constrained. A comprehensive list of the resulting *in silico* primer-GenBank matches has been provided to facilitate species-level searches across Chlorophyta, Ochrophyta and Rhodophyta, streamlining the process for researchers to identify potentially suitable primers that were developed in this study (Suppl. material 1: table S10). The findings suggest that the primers reported here are highly effective for both barcoding and metabarcoding applications.

Corallina conundrums

As a genus whose species typically lack species-specific morpho-anatomical characters, *Corallina* species delimitation is heavily reliant on DNA-based methods (i.e. cryptic species). Sequencing type specimens has helped to resolve the application of names for some species and consequent synonymy following nomenclatural priority (e.g. Brodie et al. (2013); Bustamante et al. (2019); Schipper et al. (2023); Nelson et al. (2025); Wade et al. (2025)). However, defining a species based on molecular findings has its own caveats, as using a single gene, or portion thereof, for species recognition can be misleading when genetic divergence of the targeted gene regions is low (Saunders and McDevit 2012; Bartolo et al. 2020; Calderon et al. 2021). Species identification is still possible using a single informative gene region, as demonstrated here for both *rbc*L and COI, each of which independently enables identification. However, it has long been advocated that, to establish robust species boundaries, multiple informative genes need to be used (Saunders and McDevit 2012; Twist et al. 2019; Bartolo et al. 2020). The direct use of SNPs has great potential use for species delimitation and was demonstrated to assist in species separation in this research by using approaches more common in population genetics (e.g. PCA; Leaché et al. (2014); Yesson et al. (2018); Hausdorf (2025)). Although limited by replicate numbers to include all species and only using SNPs from two genes, significant differences were identified amongst species.

Corallina berteroi is globally widespread, but has proven to be a highly variable species (molecular and morpho-anatomical data) and there is potential that there are multiple evolving populations, if not species (Calderon et al. 2021). This variability has been captured by analysing the SNPs and we encourage readers to explore the genetic diversity of *C. berteroi* in the interactive 3D PCA plots (available on GitHub). Calderon et al. (2021) suggested that

the evolutionary reversals (gains and losses of coralline fronds) seen in some *Corallina* taxa may "occur rapidly and reflect simple genetic changes". Whether this is captured in the genes targeted requires further research.

Calderon et al. (2021) recognised two clades of *Corallina* from New Zealand, previously identified as "*C. cæspitosa*", as *Corallina* sp. 1 and *Corallina* sp. 2. Type material of both *C. cæspitosa* and *C. ferreyræ* were placed within the *C. berteroi* clade recognised by Calderon et al. (2021). In this study, the application of names was updated (i.e. many samples from New Zealand, formerly placed in *C. cæspitosa*, were assigned to *C. berteroi*), but it is evident that "*C. cæspitosa*" has been applied incorrectly to other *Corallina* species. With the addition of *C. armata* (this study), *rbc*L indicates that one of the New Zealand specimens ("*C. cæspitosa*"; GenBank Accession KM369146) had an identical match and was misidentified (although further work would need to be conducted with an additional marker(s) for greater resolution and, thus, certainty). Regardless, what makes the match to *C. armata* interesting is that this "*C. cæspitosa*" specimen was taken from Campbell Point (Otago), which is approximately 800 km away from the *C. armata* type locality. This indicates that *C. armata* potentially has a wide distribution in New Zealand. This exact match between *C. armata* and the one "*C. cæspitosa*" specimen's sequence resulted in no significant difference being identified between the two species (pairwise PERMANOVA). Furthermore, given the dispersed placement of the other "*C. cæspitosa*" specimens in the PCA plot, there is a need for more investigation into *Corallina* within New Zealand. *Corallina berteroi* has priority over *C. cæspitosa*, but re-assessing specimens identified as "*C. cæspitosa*" needs to be done cautiously given the ongoing taxonomic challenges (i.e. not all specimens formerly identified as *C. cæspitosa* are *C. berteroi*). The "*C. cæspitosa*" specimens from New Zealand also have relatively high SNP variation amongst themselves and, thus, this may indicate undiscovered *Corallina* species within New Zealand. These findings open the door for interesting research to investigate whether the results align with geography or morphology within and beyond New Zealand.

While there is a lot more that can be mentioned about the complexity of *Corallina* and the unresolved phylogeny, this study has presented a way in which to delve into species clustering spatially (combining tools used in phylogenetics with population genetics). Often *Corallina* studies indicate that there are incongruencies in placement of species amongst the gene regions targeted (e.g. Schipper et al. (2023); Wade et al. (2023)). The method presented here used SNPs from two informative gene regions to identify nearest neighbours. A limitation was the inability to combine SNPs from COI and *rbc*L to compare all *Corallina* species because of restricted voucher specimens sequenced for both regions. The number of SNPs per primer set was higher than anticipated. Upon further investigation, there appears to be a few incorrectly called SNPs unique to specific individuals inflating the diversity, but this is something that cannot be resolved easily, given the need for the raw sequence file to check the quality of the call. This is one major benefit of using next-generation/third-generation sequencing platforms regarding sequencing accuracy for the entire sequence length (> 99%), reducing trimming often needed in Sanger sequencing (Wenger et al. 2019).

Reflection on short-read primer performance

As expected, there is a general increase in taxonomic assignment capabilities with the increase in the number of SNPs. This made COI a high performer given the 85 SNPs detected in comparison to the 46 detected in Rho5-7 (highest performing in terms of SNP count) for *Corallina*. Thus, COI remains an important gene to target in algae studies (if possible) given its greater resolution for species-level identifications (albeit more limited than *rbc*L in terms of database entries). Regardless, our results suggest that the short region targeted by Rho5-7 is the most suited for species-level identification of the three Rhodophyta primer sets tested as it provided the greatest resolution amongst the closely-related species, which is further supported by the *in silico* results (Fig. 10). Moreover, this middle portion of the gene region has a wealth of sequence entries in GenBank, which aids in taxonomic assignments. Considering that the overlap seen between "*C. cæspitosa*" and *C. armata* is a taxonomic mislabelling issue, the only overlap amongst *Corallina* species for Rho5-7 was seen for *C. berteroi* and *C. yendoi*. This overlap indicates that this portion of the gene region does not have enough resolution to separate these two species. An overlap was also seen between *C. berteroi* and *C. hombronii* for Rho1-3. Thus, using both Rho1-3 and Rho5-7 or longer portions of the *rbc*L gene region is needed for separating these closely-related species (Fig. 5). Rho7-8 was found to be the least informative for separating *Corallina* species and, thus, the combination of using Rho1-3 and Rho5-7 would likely provide a comprehensive species list for environmental studies. Although based on the *Corallina* alignment, the Rho1-3 region would likely not perform as well for taxonomic assignment as coverage on the 5' is more limited given the sequence entries available and this is likely the same for entries of other red algae species given the trimming needing to take place for high quality Sanger sequences. Thus, including Rho7-8 would be beneficial to increase possible species identification. Although this *in vitro* assessment of the Rhodophyta short-read primers is based on *Corallina*, these findings are an ideal starting point for other researchers due to the robust *in silico* results. If the primers are used together in a multiplex, the extension time would need to be adjusted, based on the master mix/Taq to ensure that, for example, Rho5-8 does not amplify if wanting to ensure short-reads are predominantly generated in the PCR. Furthermore, the PCR products indicate differing amplification efficiency for *Corallina* across the primer sets, which would influence proportion of reads per primer set if done in a multiplex. Ultimately, targeting multiple genes (e.g. *rbc*L and COI) and the more divergent regions within, that have high species coverage in databases, provides increased probability of taxonomic assignments, which are critical to biodiversity studies (Fig. 7; Ficetola et al. (2010)).

Interestingly, there was consistent amplification of *C. armata* and *C. berteroi* (although varying in relative read abundance) across the samples using Och6-9 primer set, yet Och6 differs at three base positions when matched to the sequenced *Corallina* species (primer base position three, four and 19; Och9 is an exact match). However, the ASV is a perfect match to the Sanger sequence for the region and, thus, did not affect amplification accuracy, but the mismatch at position 19 (3' end; penultimate) likely reduced efficiency (Bru et al. 2008; Huang et al. 2024). Nevertheless, if the goal of sequencing is to use

overlapping amplicon sequences for the alignment of one long sequence, then the efficiency of amplification is less relevant. For example, with the amplification of Och6-9, it would have been possible to generate a 612-base sequence for the *Corallina* samples using Rho5-7, Rho7-8 and Och6-9. This could of course also be adjusted as needed, based on primer selection and maximum amplicon size appropriate for sequencing (as per the mix-and-match approach taken in designing the modular primer suites).

The sensitive nature of next-generation sequencing and, thus, detecting taxa other than *Corallina* was expected given the universal primer designs. For the Rhodophyta primer sets, Rhodophyta taxa were typically amplified. The majority of the resulting ASVs assigned as *Corallina* had the highest relative abundance (Fig. 3), as *Corallina* DNA would naturally have been dominant in the extraction. On the other hand, the short-read Ochrophyta primer sets largely reflect the desired primer design so that amplification would be biased towards Ochrophyta taxa (Fig. 3). More Rhodophyta taxa could have been retained if first removing *Corallina* assigned ASVs and then filtering for RRA (less so for Ochrophyta given the lower *Corallina* RRA percentages). The filtering parameters were set so that ASVs that had at least 1% RRA were retained, but naturally this would remove many ASVs as the target ASVs would dominate in read abundance. Removing *Corallina* ASVs, would tell a more inclusive story about the local area (though likely to be lower in read numbers). In general, the detection of taxa other than *Corallina* is likely associated with the *Corallina* specimens themselves as epiphytes and/or endophytes or as DNA remnants from the local environment. Regardless, the detection of these taxa provide confidence in the inclusivity of the primer design (i.e. universal primers), while highlighting the need to increase local voucher specimen reference sequences.

Primer development and DNA databases reflections

The novelty of primer development is long gone, with several primer design softwares available (e.g. PrimerMiner, Elbrecht and Leese (2017)) and more recent software integrating artificial intelligence (AI) and machine learning (e.g. AutoPVPrimer, Ghorbani et al. (2025)). Some journals restrict the publishing of new primers as stand-alone work, while there are also many articles published utilising their newly-developed primers to investigate ecological questions without providing sufficient design details. Yet, the need for robust primers is still of great importance to the scientific community as PCR-based studies still play a fundamental role in the genomic era, reflected through their versatility in various applications (e.g. health – Dutta et al. (2022), food and wildlife forensics – Staats et al. (2016), plant and animal biodiversity surveys – Deiner et al. (2017)). Furthermore, with databases increasing with reference entries (Sayers et al. 2024), it is possible to design versatile primers that are of great benefit to the scientific community.

Entries for New Zealand macroalgae species within GenBank are known to be relatively poor in comparison to the number of endemic species known (Preuss and Zuccarello (2024); although this problem is not necessarily specific to New Zealand – Bartolo et al. (2020)). The number of entries is also far larger for *rbc*L than other gene regions across Chlorophyta, Ochrophyta and Rhodophyta (Fig. 7). Resolving entry limitations is critical as molecular methods

become more affordable for ecological studies (Keck et al. 2023). Whole genome sequencing is one alternative, but given plants are notorious for having repeat-rich regions and can also be polyploid, both influencing genome size, their assemblies are often more complex (Claros et al. 2012; Lee et al. 2018; Wang et al. 2020; Kim et al. 2022). The advent of long-read sequencers, such as those produced by Oxford Nanopore Technology and PacBio, have the capability to assist in circumventing repeat region assembly problems (Gueidan et al. 2019; Sauvage et al. 2019; Wang et al. 2021). This advancement in technology will likely increase the number of algal genomes available in the future, but, presently, there are still far fewer genomes published than their terrestrial counterparts. For example, Magnoliopsida (class; flowering plants) has over 2,400 reference genomes available in GenBank in comparison to the 280 available for Chlorophyta, Ochrophyta and Rhodophyta combined (accessed 30/07/2025; currently there is no complete genome of any *Corallina* species). Thus, while whole genomes would be the gold standard by providing the genes typically used for species-level identification in one go, adding sequences of PCR targeted genes known to resolve species identification for algae is still of immense immediate benefit to the algae community, given the low cost and quick turnaround time from DNA extraction to sequence upload on public databases (Twyford et al. 2024). This targeted method also aids in a relatively cost-effective means to grow entries for single species from different regions within countries to increase confidence in species identification when there is substantial genetic variation within a gene, which can result in a correct species match, but below accepted species confidence thresholds (van der Reis et al. 2025).

The *rbc*L primer suites developed for macroalgae in this study were designed to help facilitate database reference entries for New Zealand algae by creating a modular design to increase the number of species they could be applied to. *In silico* testing on the entire GenBank database also indicated promise for international algal community use (Suppl. material 1: table S10). The comprehensive primer-species list is also exact primer-sequence matching and, thus, the primers may still amplify other species depending on the number and position of the mismatches (Bru et al. 2008; Stadhouders et al. 2010; Lefever et al. 2013). Furthermore, out of the ~ 53,000 GenBank entries the primers were run against, there were ~ 500 that did not match a single primer and ~ 3,000 matched up to two and it is also possible that some entries may be a single short region limiting primer matching from taking place. While this study focused on short-read primer sets, it provides a comprehensive list of new *rbc*L primers that can be used for standard DNA barcoding that spans a large portion the *rbc*L gene region (as shown in this study by using Rho1-8 to aid in species characterisation). Presented are all the primers designed, even if they did not match our primer criteria (e.g. low annealing temperature) as they may still be useful if slightly modified (Suppl. material 1: table S9). The primers can be paired up as needed, as a species may match a forward primer designed for Chlorophyta and a reverse for Rhodophyta (for example; Suppl. material 1: table S10), including creating combinations to be used in a multiplex tiling approach. Rigorous *in vitro* testing of the *rbc*L primer performance on a plethora of algal species was outside the scope of this study, but the *in silico* and the *Corallina* DNA metabarcoding results suggests that these primers provide a very promising basis for future endeavours. We have started using some of

these primers beyond this case study (data not shown) and they are proving of great use in our algae-related research.

DNA metabarcoding is a popular modern technique typically used in molecular ecology to investigate biodiversity (e.g. Erlania et al. (2023); van der Reis et al. (2023); Schulte et al. (2024)). The short-read *rbc*L primer sets developed here will also be beneficial in environmental DNA metabarcoding studies, as environmental DNA is typically in a state of degradation (e.g. Collins et al. (2018); Wood et al. (2020)), thus making it more difficult to successfully amplify long gene regions when DNA is fragmented. Briefly, DNA metabarcoding primarily uses primers that target a single or multiple gene regions, or portions thereof, that hold enough diversity to allow for species identification (Taberlet et al. 2012). Primers commonly used are 'universal barcodes' that capture a broad range of taxa (e.g. Eukaryota; 18S V4 rRNA – Stoeck et al. (2010), COI – Leray et al. (2013)) or more specific groups of taxa (e.g. plants; 23S V5 rRNA – Sherwood and Presting (2007), *tuf*A – Sauvage et al. (2016)). In general, there has been less focus on creating modular universal primer suites spanning entire gene regions across a wide swath of diversity (e.g. Saunders and Moore (2013)), as it is more common for primers to be collated from literature and then selected as appropriate. However, modular universal primer suites are attractive as they help bridge the gap between barcoding and metabarcoding research, i.e. the ability to mix-and-match primer sets to target various segments and/or lengths of a gene region. Thus, this approach has the additional benefit of mitigating possible amplification issues due to primer biases, naturally improving the fidelity of DNA amplification and subsequent sequencing. Furthermore, for DNA metabarcoding biodiversity studies, choosing from a universal modular primer set allows one to capitalise on targeting the portion of a gene highly represented in databases and, thus, increasing the probability of taxonomic assignment.

While the selected Ochrophyta and Rhodophyta primers proved successful, the Chlorophyta primers are yet to be tested, but we have provided them regardless. Previous attempts to target *tuf*A from environmental samples using universal primers failed (Sauvage et al. 2016; van der Reis et al. 2023) and, thus, attempts have been made to provide alternatives given the number of Chlorophyta entries for *rbc*L in GenBank. Of note is Rho5 (CARCCWTTYATGCGYTGGAR) and Chl4 (CARCCDTTYATGCGTTGG) (Suppl. material 1: table S9) having very similar primer sequences, a version of these having been published previously as GrbcLFi (TCTCARCCWTTYATGCGTTGG; Saunders and Kucera (2010)). The latter has commonly been used as the forward primer in Chlorophyta barcoding PCR reactions; thus, given the primer sequence similarity, it is likely the reverse primers determine the specificity of targeting red and green algae (Saunders and Kucera 2010; Bartolo et al. 2020; Twyford et al. 2024). A more appropriate forward primer to target red and green algae may be to combine the primers to become CARCCDTTYATGCGYTGG and then a future goal would be to identify appropriate reverse primers for targeting red and green algae in a single reaction.

Conclusions

Species within the genus *Corallina* present significant challenges for traditional taxonomy due to the absence of distinct morpho-anatomical traits at

the species level. DNA sequencing has become instrumental in resolving species identities, enabling more rigorous and accurate application of taxonomic names. While genomics is playing an increasing role in systematics and taxonomy, its implementation remains complex and standard DNA barcoding continues to play a vital role in delineating species boundaries. The modular universal *rbc*L primer suites, developed for Ochrophyta and Rhodophyta, have proven effective for both barcoding and metabarcoding applications for *Corallina* and the results show strong potential in their application to other species. This study underscores the importance of employing multiple genetic markers, as each reveals different evolutionary histories that influence species clustering. More specifically, when using short-read approaches, *rbc*L can limit species resolution depending on the region of the gene targeted for particular species, whereas this was not seen for COI. Overall, the integration of SNP data from diverse markers provides a more comprehensive and informative framework, demonstrating the value of population genetic approaches in modern taxonomic research. To build on the concluding statement of Baba et al. (1988), "unravelling the systematics and nomenclature of [*Corallina*] is an ongoing project", we recognise that significant progress has been made and the resolution of this longstanding taxonomic challenge appears increasingly attainable. With continued research and collaboration, we remain confident that a comprehensive and stable classification of *Corallina* will be realised.

Acknowledgements

Bridget Hatton and Jean Claude Stahl (WELT) are thanked for their assistance with registration and imaging of specimens, staff of Trinity College Dublin and Natural History Museum London for access to collections. Viviana Peña for sharing results of attempts to sequence *C. armata* material, housed in the Natural History Museum. The Royal Society Te Apārangi is thanked for support to WAN through the Charles Fleming Award. The authors wish to acknowledge the use of New Zealand eScience Infrastructure (NeSI) high performance computing facilities, consulting support and/or training services as part of this research. New Zealand's national facilities are provided by NeSI and funded jointly by NeSI's collaborator institutions and through the Ministry of Business, Innovation & Employment's Research Infrastructure programme. URL https://www.nesi.org.nz.

Additional information

Conflict of interest

The authors have declared that no competing interests exist.

Ethical statement

No ethical statement was reported.

Use of AI

AI was used minimally, typically to improve R code and the clarity of some sentences, which in all cases were originally fully written by the lead author.

Funding

This work was supported by the Royal Society of New Zealand Marsden (grant number 23-UoA-126).

Author contributions

Funding acquisition and project administration for the study was equally led by MAS and WAN (project supervisors), with ALVDR supporting. ALVDR, MAS and WAN contributed to the conception of the study and provision of materials, resources and tools. WAN collected the specimens. RMW provided (at the time) unpublished sequencing data for various *Corallina* species. ALVDR developed the methodology, performed the experiments, conducted the formal analyses and validation. ALVDR also wrote all code and undertook data curation. ALVDR prepared the original draft and all visualisations (with support from WAN). All authors contributed to the development of the final version of the manuscript.

Author ORCIDs

Aimee L. van der Reis https://orcid.org/0000-0003-3054-284X
Mary A. Sewell https://orcid.org/0000-0002-1595-7951
Rachael M. Wade https://orcid.org/0000-0001-7595-2311
Wendy A. Nelson https://orcid.org/0000-0003-0014-5560

Data availability

The raw Illumina MiSeq data are available on GenBank's Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra/PRJNA1390827). The resulting barcodes are available on GenBank under accessions PX829136–PX829140 for rbcL and PX829141–PX829145 for COI. All the resulting data and scripts necessary to reproduce the analyses reported in this study can be accessed through Figshare (https://doi.org/10.17608/k6.auckland.29938859). For ease of viewing, the Bash scripts and R code are also available on GitHub (https://github.com/AvdReis/Marsden_CCA).

References

Adams NM (1994) Seaweeds of New Zealand: An illustrated guide. Canterbury University Press, Christchurch, New Zealand, 360 pp.

Akita S, Murasawa H, Kondo M, Takano Y, Kawakami Y, Nagai S, Fujita D (2020) DNA metabarcoding analysis of macroalgal seed banks on shell surface of the limpet *Niveotectura pallida*. European Journal of Phycology 55(4): 467–477. https://doi.org/10.1080/09670262.2020.1750056

Baba M, Johansen HW, Masaki T (1988) The segregation of three species of *Corallina* (Corallinales, Rhodophyta) based on morphology and seasonality in northern Japan. Botanica Marina 31(1): 15–22. https://doi.org/10.1515/botm.1988.31.1.15

Bartolo AG, Zammit G, Peters AF, Küpper FC (2020) The current state of DNA barcoding of macroalgae in the Mediterranean Sea: Presently lacking but urgently required. Botanica Marina 63(3): 253–272. https://doi.org/10.1515/bot-2019-0041

Basu C (2015) PCR Primer Design. Springer, New York, XIII, 276 pp. https://doi.org/10.1007/978-1-0716-1799-1

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener

C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, Kreps J, Langille MGI, Lee J, Ley R, Liu Y-X, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson MS II, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso JG (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology 37(8): 852–857. https://doi.org/10.1038/s41587-019-0209-9

Borer G, Monteiro C, Lima FP, Martins FMS (2025) Performance of DNA metabarcoding vs. morphological methods for assessing intertidal turf and foliose algae diversity. Molecular Ecology Resources 25(7): e14115. https://doi.org/10.1111/1755-0998.14115

Brodie J, Walker RH, Williamson C, Irvine LM (2013) Epitypification and redescription of *Corallina officinalis* L., the type of the genus, and *C. elongata* Ellis et Solander (Corallinales, Rhodophyta). Cryptogamie, Algologie 34: 48–56 https://doi.org/10.7872/crya.v34.iss1.2013.49.

Brodie J, Melbourne L, Mrowicki RJ, Brickle P, Russell S, Scott S (2021) *Corallina* (Corallinales, Rhodophyta) from Tristan da Cunha and the Falkland Islands: Implications for South Atlantic biogeography. European Journal of Phycology 56(1): 94–104. https://doi.org/10.1080/09670262.2020.1780635

Bru D, Martin-Laurent F, Philippot L (2008) Quantification of the detrimental effect of a single primer-template mismatch by real-time PCR using the 16S rRNA gene as an example. Applied and Environmental Microbiology 74(5): 1660–1663. https://doi.org/10.1128/AEM.02403-07

Bustamante DE, Calderon MS, Hughey JR (2019) Conspecificity of the Peruvian *Corallina ferreyræ* with *C. cæspitosa* (Corallinaceae, Rhodophyta) inferred from genomic analysis of the type specimen. Mitochondrial DNA. Part B, Resources 4(1): 1285–1286. https://doi.org/10.1080/23802359.2019.1591203

Calderon MS, Bustamante DE, Gabrielson PW, Martone PT, Hind KR, Schipper SR, Mansilla A (2021) Type specimen sequencing, multilocus analyses, and species delimitation methods recognize the cosmopolitan *Corallina berteroi* and establish the northern Japanese *C. yendoi* sp. nov. (Corallinaceae, Rhodophyta). Journal of Phycology 57(5): 1659–1672. https://doi.org/10.1111/jpy.13202

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods 13(7): 581–583. https://doi.org/10.1038/nmeth.3869

Callahan BJ, McMurdie PJ, Holmes SP (2017) Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. The ISME Journal 11(12): 2639–2643. https://doi.org/10.1038/ismej.2017.119

CBOL PWG Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, van der Bank M, Chase MW, Cowan RS, Erickson DL, Fazekas AJ, Graham SW, James KE, Kim K-J, Kress WJ, Schneider H, van AlphenStahl J, Barrett SCH, van den Berg C,

Bogarin D, Burgess KS, Cameron KM, Carine M, Chacón J, Clark A, Clarkson JJ, Conrad F, Devey DS, Ford CS, Hedderson TAJ, Hollingsworth ML, Husband BC, Kelly LJ, Kesanakurti PR, Kim JS, Kim Y-D, Lahaye R, Lee H-L, Long DG, Madriñán S, Maurin O, Meusnier I, Newmaster SG, Park C-W, Percy DM, Petersen G, Richardson JE, Salazar GA, Savolainen V, Seberg O, Wilkinson MJ, Yi D-K, Little DP (2009) A DNA barcode for land plants. Proceedings of the National Academy of Sciences 106: 12794–12797. https://doi.org/10.1073/pnas.0905845106

Claros MG, Bautista R, Guerrero-Fernández D, Benzerki H, Seoane P, Fernández-Pozo N (2012) Why assembling plant genome sequences is so challenging. Biology 1(2): 439–459. https://doi.org/10.3390/biology1020439

Collins RA, Wangensteen OS, O'Gorman EJ, Mariani S, Sims DW, Genner MJ (2018) Persistence of environmental DNA in marine systems. Communications Biology 1(1): 185. https://doi.org/10.1038/s42003-018-0192-6

Daugbjerg N, Andersen RA (1997) Phylogenetic analyses of the *rbc*L sequences from haptophytes and heterokont algae suggest their chloroplasts are unrelated. Molecular Biology and Evolution 14(12): 1242–1251. https://doi.org/10.1093/oxfordjournals.molbev.a025733

Deiner K, Bik HM, Mächler E, Seymour M, Lacoursière Roussel A, Altermatt F, Creer S, Bista I, Lodge DM, Vere N, Pfrender ME, Bernatchez L (2017) Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. Molecular Ecology 26(21): 5872–5895. https://doi.org/10.1111/mec.14350

Dutta D, Naiyer S, Mansuri S, Soni N, Singh V, Bhat KH, Singh N, Arora G, Mansuri MS (2022) COVID-19 diagnosis: A comprehensive review of the RT-qPCR method for detection of SARS-CoV-2. Diagnostics 12(6): e1503. https://doi.org/10.3390/diagnostics12061503

Elbrecht V, Leese F (2017) PrimerMiner: An r package for development and in silico validation of DNA metabarcoding primers. Methods in Ecology and Evolution 8(5): 622–626. https://doi.org/10.1111/2041-210X.12687

Erlania BA, Bellgrove A, Macreadie PI, Young MA, Holland OJ, Clark Z, Ierodiaconou D, Carvalho RC, Kennedy D, Miller AD (2023) Patterns and drivers of macroalgal 'blue carbon' transport and deposition in near-shore coastal environments. The Science of the Total Environment 890: e164430. https://doi.org/10.1016/j.scitotenv.2023.164430

Evans KM, Wortley AH, Mann DG (2007) An assessment of potential diatom "barcode" genes (cox1, rbcL, 18S and ITS rDNA) and their effectiveness in determining relationships in *Sellaphora* (Bacillariophyta). Protist 158(3): 349–364. https://doi.org/10.1016/j.protis.2007.04.001

Ficetola GF, Coissac E, Zundel S, Riaz T, Shehzad W, Bessière J, Taberlet P, Pompanon F (2010) An In silico approach for the evaluation of DNA barcodes. BMC Genomics 11(1): e434. https://doi.org/10.1186/1471-2164-11-434

Freshwater DW, Fredericq S, Butler BS, Hommersand MH, Chase MW (1994) A gene phylogeny of the red algae (Rhodophyta) based on plastid *rbc*L. Proceedings of the National Academy of Sciences of the United States of America 91(15): 7281–7285. https://doi.org/10.1073/pnas.91.15.7281

Freshwater D, Tudor K, O'Shaughnessy K, Wysor B (2010) DNA barcoding in the red algal order Gelidiales: Comparison of COI with rbcL and verification of the "barcoding gap". Cryptogamie. Algologie 31: 435–449.

Geller J, Meyer C, Parker M, Hawk H (2013) Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in

all-taxa biotic surveys. Molecular Ecology Resources 13(5): 851–861. https://doi.org/10.1111/1755-0998.12138

Ghorbani A, Rostami M, Ashrafi-Dehkordi E, Guzzi PH (2025) AutoPVPrimer: A comprehensive AI-Enhanced pipeline for efficient plant virus primer design and assessment. PLoS ONE 20(1): e0317918. https://doi.org/10.1371/journal.pone.0317918

Gueidan C, Elix JA, McCarthy PM, Roux C, Mallen-Cooper M, Kantvilas G (2019) PacBio amplicon sequencing for metabarcoding of mixed DNA samples from lichen herbarium specimens. MycoKeys 53: 73–91. https://doi.org/10.3897/mycokeys.53.34761

Guiry MD (2024) How many species of algae are there? A reprise. Four kingdoms, 14 phyla, 63 classes and still growing. Journal of Phycology 60(2): 214–228. https://doi.org/10.1111/jpy.13431

Guiry MD, Andersen RA, Moestrup Ø (2023) Validation of the name Heterokontophyta. Notulae Algarum 297: 1–5.

Hadi SI, Santana H, Brunale PP, Gomes TG, Oliveira MD, Matthiensen A, Oliveira ME, Silva FC, Brasil BS (2016) DNA barcoding green microalgae isolated from neotropical inland waters. PLoS ONE 11(2): e0149284. https://doi.org/10.1371/journal.pone.0149284

Harvey WH (1849) *Nereis australis* or algae of the southern ocean: Being figures and descriptions of marine plants collected on the shores of the Cape of Good Hope the extra-tropical Australian colonies Tasmania New Zealand and the Antarctic regions; deposited in the Herbarium of the Dublin University. Reeve Brothers, London (Part 2): 65–124.

Hausdorf B (2025) Species delimitation using genomic data: Options and limitations. Molecular Ecology 34(8): e17717. https://doi.org/10.1111/mec.17717

Hind KR, Saunders GW (2013) A molecular phylogenetic study of the tribe Corallineae (Corallinales, Rhodophyta) with an assessment of genus-level taxonomic features and descriptions of novel genera. Journal of Phycology 49(1): 103–114. https://doi.org/10.1111/jpy.12019

Hind KR, Gabrielson PW, Lindstrom SC, Martone PT (2014) Misleading morphologies and the importance of sequencing type specimens for resolving coralline taxonomy (Corallinales, Rhodophyta): *Pachyarthron cretaceum* is *Corallina officinalis*. Journal of Phycology 50(4): 760–764. https://doi.org/10.1111/jpy.12205

Huang K, Zhang J, Li J, Qiu H, Wei L, Yang Y, Wang C (2024) Exploring the impact of primer–template mismatches on PCR performance of DNA polymerases varying in proofreading activity. Genes 15(2): e215. https://doi.org/10.3390/genes15020215

Illumina (2013) 16S metagenomic sequencing library preparation: Preparing 16S ribosomal RNA gene amplicons for the Illumina MiSeq system. [Report no. 15044223 Rev. B]. https://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf

Kahle D, Wickham H (2013) ggmap: Spatial visualization with ggplot2. The R Journal 5(1): 144–161. https://doi.org/10.32614/RJ-2013-014

Kans J (2013) Entrez Direct: E-utilities on the Unix Command Line. Entrez Programming Utilities Help (Updated 10 Oct 2024). National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/books/NBK179288/

Katoh K, Misawa K (2002) MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30(14): 3059–3066. https://doi.org/10.1093/nar/gkf436

Keck F, Couton M, Altermatt F (2023) Navigating the seven challenges of taxonomic reference databases in metabarcoding analyses. Molecular Ecology Resources 23(4): 742–755. https://doi.org/10.1111/1755-0998.13746

Kezlya E, Tseplik N, Kulikovskiy M (2023) Genetic markers for metabarcoding of freshwater microalgae. Biology 12(7): e1038. https://doi.org/10.3390/biology12071038 [Review]

Kim D, Lee J, Cho CH, Kim EJ, Bhattacharya D, Yoon HS (2022) Group II intron and repeat-rich red algal mitochondrial genomes demonstrate the dynamic recent history of autocatalytic RNAs. BMC Biology 20(1): 1–2. https://doi.org/10.1186/s12915-021-01200-3

Kress WJ, Erickson DL (2012) DNA barcodes: Methods and protocols. Methods in Molecular Biology 858: 3–8. https://doi.org/10.1007/978-1-61779-591-6_1

Leaché AD, Fujita MK, Minin VN, Bouckaert RR (2014) Species delimitation using genome-wide SNP data. Systematic Biology 63(4): 534–542. https://doi.org/10.1093/sysbio/syu018

Lee J, Yang EC, Graf L, Yang JH, Qiu H, Zelzion U, Chan CX, Stephens TG, Weber APM, Boo GH, Boo SM, Kim KM, Shin Y, Jung M, Lee SJ, Yim H-S, Lee J-H, Bhattacharya D, Yoon HS (2018) Analysis of the draft genome of the red seaweed *Gracilariopsis chorda* provides insights into genome size evolution in Rhodophyta. Molecular Biology and Evolution 35(8): 1869–1886. https://doi.org/10.1093/molbev/msy081

Lefever S, Pattyn F, Hellemans J, Vandesompele J (2013) Single-nucleotide polymorphisms and other mismatches reduce performance of quantitative PCR assays. Clinical Chemistry 59(10): 1470–1480. https://doi.org/10.1373/clinchem.2013.203653

Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, Boehm JT, Machida RJ (2013) A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: Application for characterizing coral reef fish gut contents. Frontiers in Zoology 10(1): 1–34. https://doi.org/10.1186/1742-9994-10-34

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal 17(1): 1–3. https://doi.org/10.14806/ej.17.1.200

Morgulis A, Coulouris G, Raytselis Y, Madden TL, Agarwala R, Schaffer AA (2008) Database indexing for production MegaBLAST searches. Bioinformatics 24(16): 1757–1764. https://doi.org/10.1093/bioinformatics/btn322

Mshiywa FM, Edwards S, Bradley G (2024) Rhodophyta DNA barcoding: Ribulose-1, 5-bisphosphate carboxylase gene and novel universal primers. International Journal of Molecular Sciences 25(1): 1–58. https://doi.org/10.3390/ijms25010058

Nelson W, Neill K, D'Archino R (2023) Chapter 3. Kingdom Chromista, phylum Ochrophyta, and kingdom Plantae, phyla Prasinodermatophyta, Chlorophyta, and Rhodophyta (macroalgae). In: Kelly M, Mills S, Terezow M, Sim-Smith C, Nelson W (Eds) The Marine Biota of Aotearoa New Zealand Updating our marine biodiversity inventory NIWA Biodiversity Memoir. NIWA, Wellington, New Zealand, 61–80. https://doi.org/10.2307/jj.6947045.11

Nelson W, Neill K, Le Gall L, Peña V, Twist B, Wade R (2025) Taxonomic revision of *Corallina hombronii*: Type sequencing of *Jania hombronii* reveals its synonymy with *Corallina chamberlainiæ* (Corallinales, Rhodophyta). Phycologia 64(3): 1–10. https://doi.org/10.1080/00318884.2025.2513128

Nozaki H, Itoh M, Sano R, Uchida H, Watanabe MM, Kuroiwa T (1995) Phylogenetic relationships within the colonial volvocales (Chlorophyta) inferred from *rbc*L gene sequence data. Journal of Phycology 31(6): 970–979. https://doi.org/10.1111/j.0022-3646.1995.00970.x

Prediger E (2013) How to design primers and probes for PCR and qPCR. Integrated DNA Technologies. [Retrieved 12 June 2025] https://www.idtdna.com/pages/education/decoded/article/designing-pcr-primers-and-probes

Preuss M, Zuccarello GC (2024) Review of available DNA sequence data: Our understanding of macroalgal diversity in New Zealand. New Zealand Journal of Botany 63(5): 2264–2274. https://doi.org/10.1080/0028825X.2024.2330634

R Core Team (2024). R: A language and environment for statistical computing (Version 4.4.1; R Studio Version 2024.09.0+375). R Foundation for Statistical Computing. https://www.R-project.org/

Robertson JM, Walsh-Weller J (1998) An introduction to PCR primer design and optimization of amplification reactions. In: Lincoln PJ, Thomson J (Eds) Forensic DNA Profiling Protocols. Humana Press, Totowa, 121–154. https://doi.org/10.1385/0-89603-443-7:121

San MH, Kawamura Y, Kimura K, Witharana EP, Shimogiri T, Fujiyoshi E, Nagano Y (2025) Gene tree conflict in chloroplast-based phylogenies in *Pyropia* species. Phycological Research 73(2): 79–91. https://doi.org/10.1111/pre.12576

Saunders GW, Kucera H (2010) An evaluation of *rbc*L, *tuf*A, UPA, LSU and ITS as DNA barcode markers for the marine green macroalgae. Cryptogamie. Algologie 31: 487–528.

Saunders GW, McDevit DC (2012) Methods for DNA barcoding photosynthetic protists emphasizing the macroalgae and diatoms. In: Kress WJ, Erickson DL (Eds) DNA Barcodes: Methods and Protocols. Humana Press, Totowa, 207–222. https://doi.org/10.1007/978-1-61779-591-6_10

Saunders GW, Moore TE (2013) Refinements for the amplification and sequencing of red algal DNA barcode and RedToL phylogenetic markers: A summary of current primers, profiles and strategies. Algae – Korean Phycological Society 28(1): 31–43. https://doi.org/10.4490/algae.2013.28.1.031

Sauvage T, Schmidt WE, Suda S, Fredericq S (2016) A metabarcoding framework for facilitated survey of endolithic phototrophs with *tuf*A. BMC Ecology 16(1): 1–8. https://doi.org/10.1186/s12898-016-0068-x

Sauvage T, Schmidt WE, Yoon HS, Paul VJ, Fredericq S (2019) Promising prospects of nanopore sequencing for algal hologenomics and structural variation discovery. BMC Genomics 20(1): e850. https://doi.org/10.1186/s12864-019-6248-2

Sayers EW, Cavanaugh M, Frisse L, Pruitt KD, Schneider VA, Underwood Beverly A, Yankie L, Karsch-Mizrachi I (2024) GenBank 2025 update. Nucleic Acids Research 53(D1): D56–D61. https://doi.org/10.1093/nar/gkae1114

Schipper P Sr, Shivak JP, Hind KR, Miller KA, Hughey JR, Gabrielson PW, Martone PT (2023) Reinstatement of *Corallina chilensis* (Corallinaceae, Rhodophyta) based on DNA sequencing of the type material collected by Darwin. Phycologia 62(3): 203–216. https://doi.org/10.1080/00318884.2022.2159198

Schulte NO, Craine JM, Leopold DR, Devitt JK, Fierer N (2024) Diversity of freshwater algal assemblages across the United States as revealed by DNA metabarcoding. Environmental DNA 6(1): e477. https://doi.org/10.1002/edn3.477

Sherwood AR, Presting GG (2007) Universal primers amplify a 23S rDNA plastid marker in eukaryotic algae and cyanobacteria. Journal of Phycology 43(3): 605–608. https://doi.org/10.1111/j.1529-8817.2007.00341.x

Staats M, Arulandhu AJ, Gravendeel B, Holst-Jensen A, Scholtens I, Peelen T, Prins TW, Kok E (2016) Advances in DNA metabarcoding for food and wildlife forensic species identification. Analytical and Bioanalytical Chemistry 408(17): 4615–4630. https://doi.org/10.1007/s00216-016-9595-8

Stadhouders R, Pas SD, Anber J, Voermans J, Mes THM, Schutten M (2010) The effect of primer-template mismatches on the detection and quantification of nucleic ac-

ids using the 5' nuclease assay. The Journal of Molecular Diagnostics: JMD 12(1): 109–117. https://doi.org/10.2353/jmoldx.2010.090035

Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9): 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stoeck T, Bass D, Nebel M, Christen R, Jones MDM, Breiner H-W, Richards TA (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Molecular Ecology 19(s1): 21–31. https://doi.org/10.1111/j.1365-294X.2009.04480.x

Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. Molecular Ecology 21(8): 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Thiers BM (2025) Index Herbariorum. [Retrieved 22 Sept 2025] https://sweetgum.nybg.org/science/ih/

Turland NJ, Wiersema JHB FR, Gandhi KN, Gravendyck J, Greuter W, Hawksworth DL, Herendeen PS, Klopper RR, Knapp S, Kusber W-H, Li D-Z, May TW, Monro AM, Prado J, Price MJ, Smith GF, Zamora Señoret JC (2025) International Code of Nomenclature for Algae, Fungi, and Plants (Madrid Code) (Vol. 162). University of Chicago Press, Chicago. https://doi.org/10.7208/chicago/9780226839479.001.0001

Twist BA, Neill KF, Bilewitch J, Jeong SY, Sutherland JE, Nelson WA (2019) High diversity of coralline algae in New Zealand revealed: Knowledge gaps and implications for future research. PLoS ONE 14(12): e0225645. https://doi.org/10.1371/journal.pone.0225645

Twyford AD, Beasley J, Barnes I, Allen H, Azzopardi F, Bell D, Blaxter ML, Broad G, Campos-Dominguez L, Choonea D, Crowley L, Cuber P, Cunliffe M, Dombrowski A, Douglas B, Forrest LL, Gaya E, Greeves C, Griffin C, Harley J, Hart ML, Holland PWH, Hollingsworth PM, Januszczak I, Jones A, Kersey P, Kilias E, Lawniczak MKN, Lewis OT, Mian S, Minotto A, Misra R, Mulhair PO, Pereira da Conceicoa L, Price BW, Salatino S, Shaw F, Sivell O, Sivess L, Uhl R, Woof K (2024) A DNA barcoding framework for taxonomic verification in the Darwin Tree of Life Project. Wellcome Open Research 9: e339. https://doi.org/10.12688/wellcomeopenres.21143.1

van der Reis A, Sewell M, Nelson W (2023) Investigating seed bank potential of crustose coralline algae using DNA metabarcoding. Journal of Phycology 60(1): 195–202. https://doi.org/10.1111/jpy.13403

van der Reis AL, Lalas C, Schuckard R, Middlemiss KL, Jeffs AG (2025) Cross-validation of diet determination methods for seabird conservation. Environmental DNA 7(4): e70153. https://doi.org/10.1002/edn3.70153

Wade RM, Gabrielson PW, Hind KR, Shivak J, Hughey JR, Ohtsu S, Baba M, Kogame K, Lindstrom SC, Miller KA, Schipper SR, Martone PT (2023) Resolving some of the earliest names for *Corallina* species (Corallinales, Rhodophyta) in the North Pacific by sequencing type specimens and describing the cryptic *C. hakodatensis* sp. nov. and *C. parva* sp. nov. Journal of Phycology 59(1): 221–235. https://doi.org/10.1111/jpy.13299

Wade RM, Ogushi R, Gabrielson PW, Hind KR, Hughey JR, Lindstrom SC, Miller KA, Martone PT (2025) Type-assisted taxonomy and biogeography of *Corallina* (Corallinales, Rhodophyta) in the eastern Pacific: *Corallina americana* sp. nov., *C. bathybentha*, *C. hommersandiorum* sp. nov., and *C. saundersii* sp. nov. Journal of Phycology 61(5): 1503–1520. https://doi.org/10.1111/jpy.70078

Wang S, Lin L, Shi Y, Qian W, Li N, Yan X, Zou H, Wu M (2020) First draft genome assembly of the seaweed *Sargassum fusiforme*. Frontiers in Genetics 11: e590065. https://doi.org/10.3389/fgene.2020.590065

Wang Y, Zhao Y, Bollas A, Wang Y, Au KF (2021) Nanopore sequencing technology, bioinformatics and applications. Nature Biotechnology 39(11): 1348–1365. https://doi.org/10.1038/s41587-021-01108-x

Wenger AM, Peluso P, Rowell WJ, Chang P-C, Hall RJ, Concepcion GT, Ebler J, Fungtammasan A, Kolesnikov A, Olson ND, Töpfer A, Alonge M, Mahmoud M, Qian Y, Chin C-S, Phillippy AM, Schatz MC, Myers G, DePristo MA, Ruan J, Marschall T, Sedlazeck FJ, Zook JM, Li H, Koren S, Carroll A, Rank DR, Hunkapiller MW (2019) Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nature Biotechnology 37(10): 1155–1162. https://doi.org/10.1038/s41587-019-0217-9

Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis (Vol. 2). Springer-Verlag, New York, 260 pp.

Winship PR (1989) An improved method for directly sequencing PCR-amplified material using dimethyl sulphoxide. Nucleic Acids Research 17(3): 1266–1266. https://doi.org/10.1093/nar/17.3.1266

Woelkerling WJ, Nelson WA (2004) A baseline summary and analysis of the taxonomic biodiversity of coralline red algae (Corallinales, Rhodophyta) recorded from the New Zealand region. Cryptogamie. Algologie 25: 39–106.

Wood SA, Biessy L, Latchford JL, Zaiko A, von Ammon U, Audrezet F, Cristescu ME, Pochon X (2020) Release and degradation of environmental DNA and RNA in a marine system. The Science of the Total Environment 704: e135314. https://doi.org/10.1016/j.scitotenv.2019.135314

Yesson C, Jackson A, Russell S, Williamson CJ, Brodie J (2018) SNPs reveal geographical population structure of *Corallina officinalis* (Corallinaceae, Rhodophyta). European Journal of Phycology 53(2): 180–188. https://doi.org/10.1080/09670262.2017.1402373

Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y (2017) ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution 8(1): 28–36. https://doi.org/10.1111/2041-210X.12628

Supplementary material 1

Supplementary tables and figures are provided online as an excel spreadsheet to facilitate a 'one-stop-shop' supplementary file approach

Authors: Aimee L. van der Reis, Mary A. Sewell, Rachael M. Wade, Wendy A. Nelson

Data type: xlsx

Explanation note: **table S1.** *Corallina armata* and *Corallina berteroi* collection locations. **table S2.** Indicates the initial *Corallina* entries pulled from GenBank for *rbc*L and COI (exception of sequence contribution from RMW - Wade 2025). **table S3.** Raw read filtering using Cutadapt, DADA2 and R. PrimerSets Rho and Och target *rbc*L and Leray targets COI. **table S4.** Amplicon sequence variant (ASV) table with assigned taxonomy using MegaBLAST and the GenBank nucleotide database. **table S5.** Final DNA sequences for *Corallina armata* (Carm1-4) and *C. berteroi* (Cber1) specimens for COI and *rbc*L (primers used Rho1-8). **table S6.** The number of *Corallina* voucher reference sequences that could be used for phylogenetic trees and PCA/nearest-neighbour plots. **table S7.** Pairwise PERMANOVA for *Corallina* species. **table S8.** New Zealand (NZ) species taken from Nelson et al. 2023 list were pulled from GenBank for nine gene regions. **table S9.** *rbc*L primers identified from the consensus sequence (90%) from the aligned New Zealand species. **table S10.** All *rbc*L primer matches for GenBank entries for Chlorophyta, Ochrophyta and Rhodophyta (pulled April 2025). **table S11.** Selected *rbc*L primer pairs were tested *in silico* for matches in the pulled GenBank database. **fig. S1.** COI heatmap representing pairwise *Corallina* species sequence clustering *de novo* based on genetic distances. **fig. S2.** *rbc*L heatmap representing pairwise *Corallina* species sequence clustering *de novo* based on genetic distances. **fig. S3.** PCA plot (top) and nearest-neighbour PCA-derived plot (bottom) for *Corallina* voucher specimens for Rho1-3 and Rho7-8. **fig. S4.** *Corallina armata* epitype: WELT A036058, New Zealand, North Island, Wairarapa, Mataikona, (-40.803 S, 176.260 E), 08 February 2024, eg. W.A.Nelson, AWA001 (Carm1 in paper). **fig. S5.** Chlorophyta GenBank entries and matches to *rbc*L primers designed in this study - showing family left y-axis, number of species and class on right y-axis. **fig. S6.** Ochrophyta GenBank entries and matches to *rbc*L primers - showing family left y-axis, number of species and class on right y-axis. **fig. S7.** Rhodophyta GenBank entries and matches to *rbc*L primers - showing family left y-axis, number of species and class on right y-axis.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/mbmg.10.173634.suppl1